

# **Cultural Relevance of Friends Sitcom through Text Analysis**

MY360 Quantitative Text Analysis

Group 2

LSE Candidate Numbers: 24155, 24980, 24246, 24692

Word count: 3927

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Distinctiveness of character speech and clarity of emotional conveyance through dialogues</b>	<b>2</b>
2.1	Character Detection . . . . .	2
2.2	Emotion Detection . . . . .	4
<b>3</b>	<b>Representation of Women Characters in Friends</b>	<b>7</b>
3.1	Data Pre-processing . . . . .	7
3.2	Exploratory Analysis of the Characters Relationships . . . . .	8
3.3	The Analysis of Female Character . . . . .	8
<b>4</b>	<b>Friends as an English Language Learning Resource</b>	<b>11</b>
4.1	The Diversity and Complexity of Language in Friends . . . . .	11
4.2	Variations in Language Diversity and Complexity Across Thematic Contexts . . . . .	12
4.2.1	Sentiment-Vocabulary analysis . . . . .	12
4.2.2	Context Dictionary Analysis . . . . .	13
<b>5</b>	<b>Unveiling Criticism: Analysis of Offensive Content in Friends</b>	<b>14</b>
5.1	General Analysis of Offensive Content . . . . .	14
5.2	Exploring Weight Criticism Towards Monica . . . . .	16
<b>6</b>	<b>Conclusion</b>	<b>17</b>
	<b>References</b>	<b>18</b>
	<b>Appendix</b>	<b>19</b>

# 1 Introduction

The popular sitcom *Friends*, which aired from 1994 to 2004, featured six main characters: Rachel Green, Ross Geller, Chandler Bing, Monica Geller, Joey Tribbiani, and Phoebe Buffay. Over the course of 10 seasons, viewers became closely familiar with each character's unique personality, quirks, and catchphrases. For over a decade, the show has regained popularity multiple times over the decades, demonstrating a widespread and lasting impact on culture, fashion trends, and role models.

*Friends*' enduring relevance (Cobb, Ewen, and Hamad (2018)) can be attributed to several key factors that this report explores. The distinctiveness of the characters and their relatable qualities have allowed viewers to form strong connections with them. The show's ability to convey emotions clearly through dialogue has also contributed to its appeal. Also the representation of female characters, particularly their growth and development over the series, has resonated with audiences and to some degree reinforce the public's impression on genders (Holbert, Shah, and Kwak (2003)). In addition, *Friends* has served as a valuable resource for English language learners due to its diverse vocabulary, idiomatic expressions, and portrayal of everyday conversational English (Larrea Espinar (2020)). Furthermore, despite criticism of the show's occasional offensive content, an analysis of these instances provides insight into the social and cultural context of the time and the evolution of public discourse on sensitive topics. This report connects underpinning factors of the show's relevance today: the personality and relevance of characters, its potential as an English language learning resource and the heavy critical discourse on mentions of offensive content.

This report aims to investigate how the distinctive characters, language, and controversial content in the sitcom *Friends* contribute to its enduring popularity and cultural relevance. The analysis is based on the script of the seasons from the R *Friends* package of EmilHvitfeldt (n.d.) and uses predominantly the *Quanteda* package for analysis. The dataset comprises transcripts from the show, organized by individual speaker utterances for each scene across all seasons and episodes. The data also includes emotion and character entity annotation for some of the utterances. These annotations have been included in separate tibbles, which can easily be joined to the main dataset when needed.

The structure of this report is the following: the first section aims to explore the extent to which predictive models can differentiate between the seemingly distinct personalities of the main characters in the show. The second section puts the primary focus on the character development of female characters, trying to identify the female characters representations in this 1990s US TV show. The third section provides a dictionary based analysis of the language complexity in the show across emotional and thematic contexts and its implications for English learning. The last section examines the validity of criticism the show has received the last couple of years due to its reported homophobic, sexist and/or fat shaming content.

## 2 Distinctiveness of character speech and clarity of emotional conveyance through dialogues

**LSE candidate number: 24155**

### 2.1 Character Detection

This section aims to answer the question how distinct are the characters in their way of speaking, and how clearly are their emotions conveyed through text.

The show's six main characters accounted for 76 percent of the series' spoken lines. Rachel had the most utterances at 9,238, followed by Ross at 9,134. Chandler, Monica, and Joey each had over 8,000 lines, while Phoebe had the least of the six with 7,416.

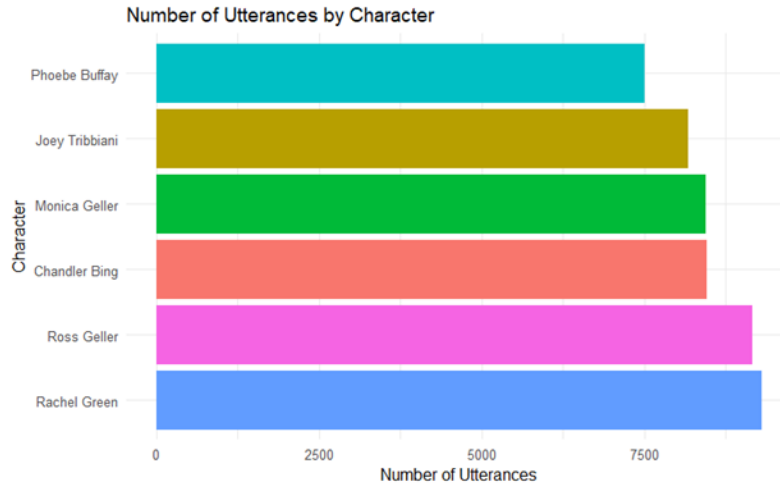


Figure 1: Character Utterances

The wordcloud visualisation shows the most frequent words of each character. Phoebe's cloud shows her quirky nature with words like "ooh", "umm", "ohh" and "yeah", while Joey's features "hey", and "umm", reflecting his more laid back personality. Rachel's cloud shows "oh", "god", "oh my", relating to her dramatic side.



Figure 2: Wordcloud for Six Characters

The Naive Bayes confusion matrix shows poor performance, with accuracy of only 27.7 % . The model struggles most with Ross, achieving only 21% sensitivity (recall). It does best on Joey and Rachel, but even their sensitivity is only around 34%. Specificity values are higher, mostly in the 80-90% range, indicating the model is better at ruling characters out than identifying them definitively. Chandler is an exception with 83% specificity.

The leading pattern is that the model frequently misclassifies other characters as Chandler. The model predicts Chandler far more than his actual prevalence. Conversely, Ross is frequently misclassified as other characters, with the model having a hard time identifying his lines. Fine-tuned distilled roberta, a transformer, achieves slightly

Predicted Label \ True Label	Chandler	Joey	Monica	Phoebe	Rachel	Ross
<b>Chandler Bing</b>	449	261	304	250	294	305
<b>Joey Tribbiani</b>	336	551	268	234	252	368
<b>Monica Geller</b>	234	198	421	228	258	260
<b>Phoebe Buffay</b>	215	191	207	381	230	212
<b>Rachel Green</b>	259	239	328	302	631	296
<b>Ross Geller</b>	183	178	182	133	189	383

Table 1: Confusion Matrix for Character Detection

better results, with accuracy of 29.7% and an F1 score of 29.5%. Compared to a naive baseline of always predicting the most frequent character (Rachel), which would get 18.2% accuracy, the transformer is an improvement, but still far from great performance. Both models seem to do best on Rachel and Joey, and worst on Ross. However, the transformer model appears to have a more balanced performance across characters, with F1-scores ranging from 0.26 to 0.34, compared to the wider range of 0.25 to 0.32 for the Naive Bayes model.

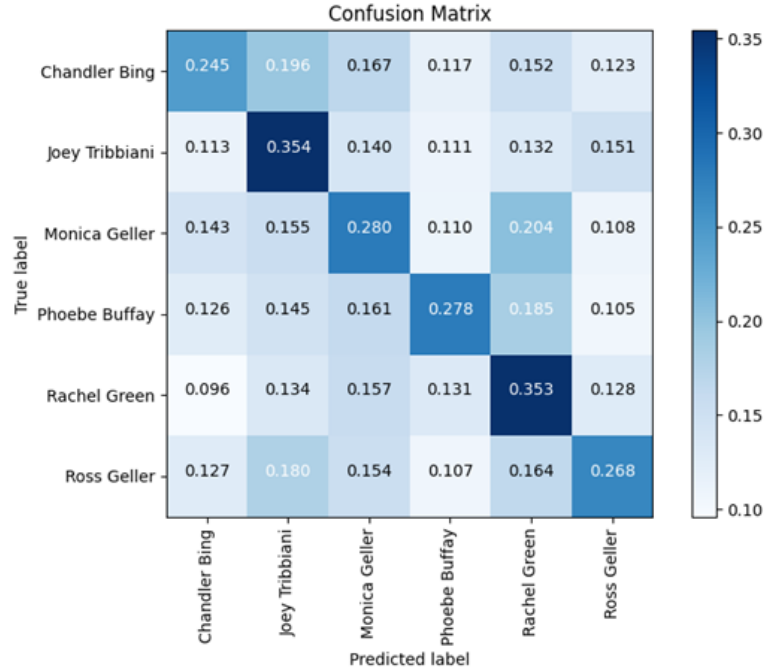


Figure 3: Distilled Roberta Model Confusion Matrix

Joey and Rachel are easier to classify, since they have many signature catchphrases and verbal tics. Joey’s ”how you doin’?” and Rachel’s ”noooo” are iconic lines that the models can latch onto. Similarly, Phoebe’s abstract non-sequiturs likely help the model identify her. In contrast, the more straitlaced Ross, Chandler and Monica have fewer unique linguistic markers.

## 2.2 Emotion Detection

We compared the performance of four models for emotion classification: a Naive Bayes model, a Multinomial Lasso model, a BERT-base-uncased model, and a fine-tuned RoBERTa model (SamLowe/roberta-base-go\_emotions).



Predicted Label \ True	Joyful	Mad	Neutral	Peaceful	Powerful	Sad	Scared
Joyful	189	46	173	54	52	28	53
Mad	52	65	69	30	35	16	34
Neutral	91	33	212	59	39	23	49
Peaceful	62	25	81	34	17	16	30
Powerful	62	24	73	18	22	15	25
Sad	48	26	80	35	26	44	32
Scared	73	34	77	21	27	26	67

Table 2: Naive Bayes Confusion Matrix

balanced classes were adjusted by calculating weights inversely proportional to the class frequencies. Both models often misclassify 'Peaceful' and 'Powerful' as 'Joyful' or 'Neutral', the two most common classes.

The BERT-base-uncased model, after 20 epochs of fine-tuning, reaches an accuracy of 37.23%, a significant improvement over the linear models, while 'Powerful' remains the hardest to classify.

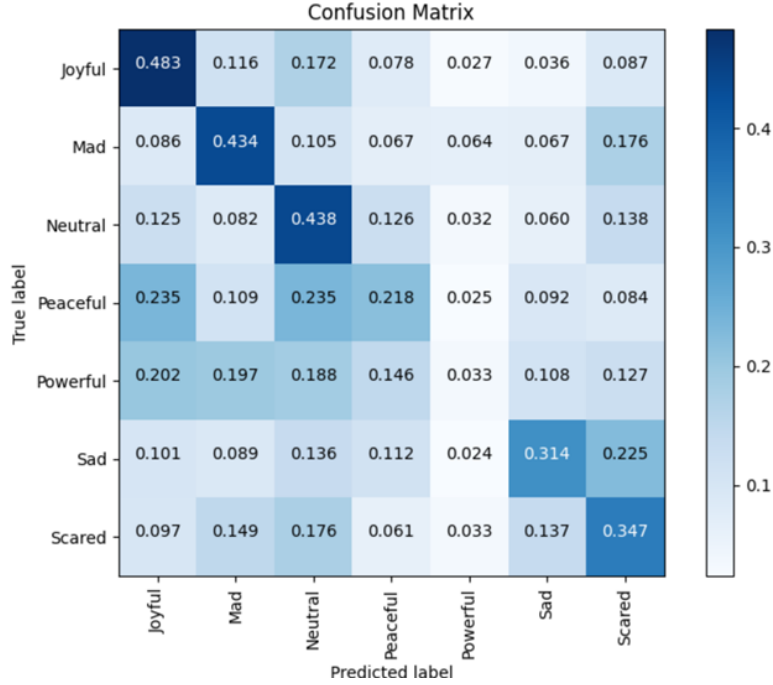


Figure 6: Bert Model Confusion Matrix

The pre-trained RoBERTa model, fine-tuned on the Friends data, achieves the best performance with an accuracy of 52.97%. The confusion matrix shows a more balanced performance across emotions, with 'Mad', 'Sad', and 'Joyful' having the highest precision and recall. 'Peaceful' remains challenging.

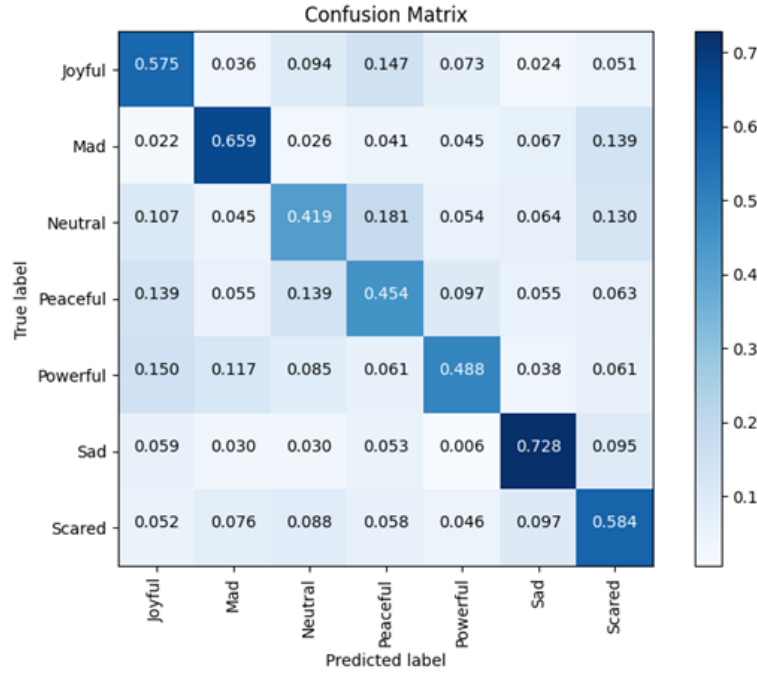


Figure 7: Roberta Model Confusion Matrix

The transformer-based models (BERT and RoBERTa) significantly outperform the simpler, linear models (Naive Bayes and Lasso) due to their ability to learn more nuanced patterns. RoBERTa’s superior performance can be attributed to its pre-training on an emotion dataset and its architectural improvements over BERT. The emotions captured by the models align with the characters’ personalities and storylines. The high frequency of ‘Joyful’ and ‘Neutral’ reflects the show’s comedic nature, while the presence of ‘Mad’ and ‘Sad’ captures the dramatic and sometimes turbulent relationships between the characters. The models’ difficulty with ‘Peaceful’ and ‘Powerful’ emotions could be due to the context-dependent ways these emotions are expressed in the show. However, even the best-performing model achieves an accuracy of only 52.97%.

### 3 Representation of Women Characters in Friends

**LSE candidate number: 24692**

The women characters representations on US TV series have changed significantly for the past few decades. Female characters in the 1980s were criticized for being only emphasized in relation to male characters (Zeisler (2016)). In contrast, the 1990s and 2000s, marked a revolution of female protagonists as they were described as more complex and morally ambivalent (Tally (2016)). The underrepresentation of females in prestigious positions is criticized to perpetuate the belief that women are intellectually inferior to men (Holbert, Shah, and Kwak (2003)). This text analysis aims to investigate how Friends shaped the female characters, taking the example from the scripts of “Rachel Green” who is one of the main characters.

#### 3.1 Data Pre-processing

The pre-processing step is conducted on the original scripts dataset friends. The subset of 6 main characters was taken as the main focus. I divided the same original scripts twice in different ways, one is divided by episodes based on the two life turning points of Rachel Green, while another is based on the gender. The dialogues contain a large amount of oral expression such as “oh” or “uhh”. These words are removed when creating the document frequency matrix (dfm) as the content of the conversation is my main interest.



### 3.2 Exploratory Analysis of the Characters Relationships

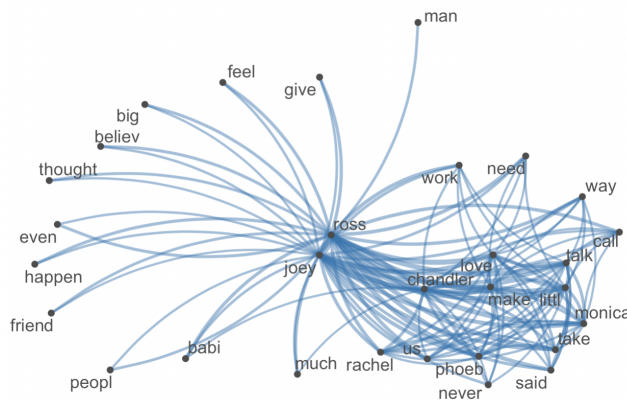


Figure 8: Network Plot of Character Relationships

Topic	Keywords	Order
Topic 4	monica, joey, talk	1
Topic 7	us, joey, call	2
Topic 2	rachel, peopl, friend	3
Topic 1	ross, said, joey	4
Topic 5	phoeb, joey, love	5
Topic 3	man, take, big	6
Topic 6	chandler, make, babi	7

Table 3: Topic Proportions Top 7

The network graph and topic model are built based on the 6 main characters' scripts documents. Each node in the network represents a word, and the lines between them show relationships or co-occurrences of these words within the scripts. Notably, names of the 6 can stand out with most connections to other words. In Particular, Ross and Joey are more connected than others. Combined with the topic model results, the male characters seem to be the center of the friend circle. In the network, "love," "friend," "feel" and "believe" are also prominent, implying these are key themes or frequently discussed topics within the texts.

### 3.3 The Analysis of Female Character

Rachel Green, growing from a cafe waitress to a fashion professional, her growth and independence made her the most welcomed character. Two life turning moments in her life: getting a professional job as an assistant in retail and becoming a leadership role in Ralph Lauren divided her scripts into three periods of subsets within unequal size. Growth of character across the periods and the gender difference in scripts will be analyzed through word cloud, dfm tf-idf weighting, sentiment analysis and dictionary methods.

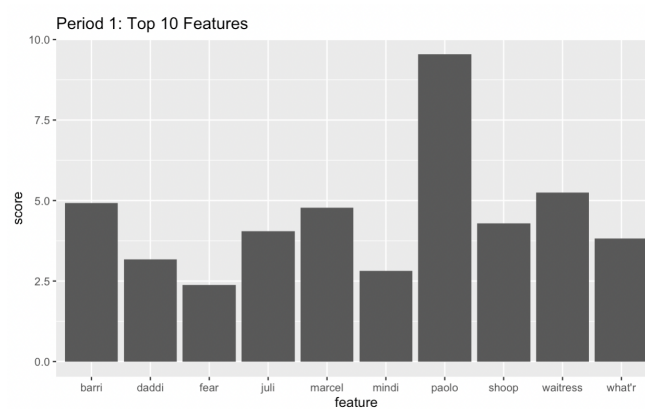


Figure 9: Top 10 features for Rachel in Period 1

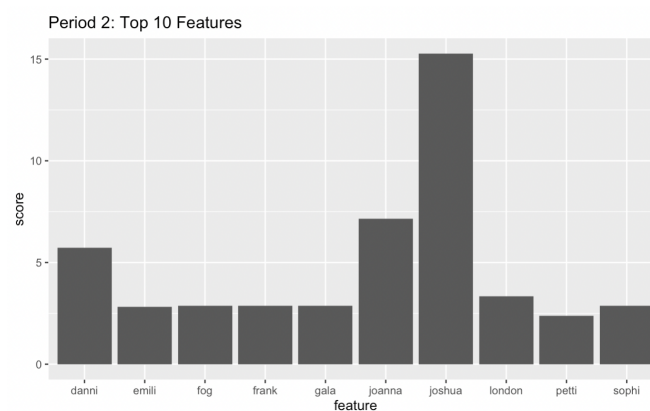


Figure 10: Top 10 features for Rachel in Period 2

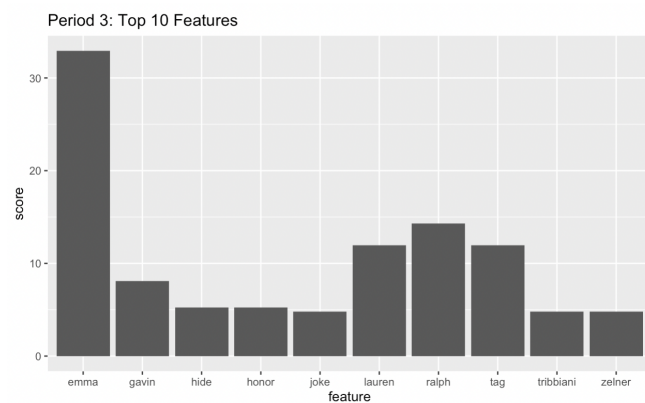


Figure 11: Top 10 features for Rachel in Period 3

Above are the top 10 features which are the most relevant words based on the dfm tf-idf scores for each period. From period1, representatives such as "daddy" and "barry\*" indicate her continued reliance on her father and Bianca. Period two emphasizes Joanna\* and Joshua\*, pivotal in her move from waitress to professional assistant, significantly advancing her career. In the final period, the focus shifts to family life, with "Emma," her daughter, as the most relevant word. Meanwhile, "Ralph Lauren" and "Zelner\*" play crucial roles in period 3. These changes underscore the significance of character relationships in narrative development and character representations.

Documents	Work_related	Money_related	Career_related
Period 1	63	15	17
Period 2	61	12	10
Period 3	154	32	60

Table 4: Rachel’s scripts development across three period

A self-defined dictionary is used to track the appearance of personal development-related words across three periods. The dfm reveals a significant increase in features counts from period one to period three, despite doubling the sample size for period three. An analysis of random instances from each period shows a progression in Rachel’s sentiment towards her job: from complaints about unemployment in period one, to neutral discussions in period two, to excitement about career progress in period three. This suggests that job-related issues increasingly dominate the narrative, likely reflecting her character development.

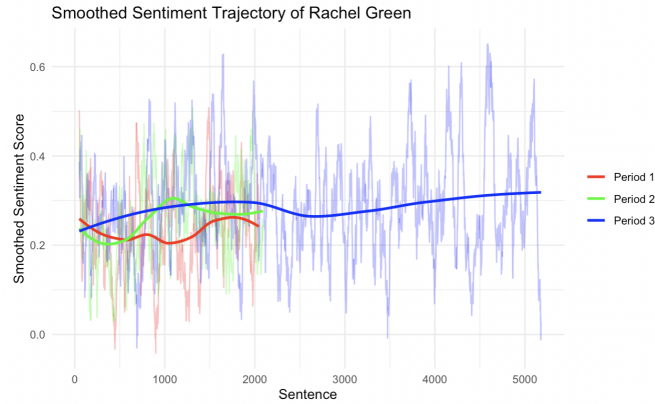


Figure 12: Rachel’s sentiment Analysis

The plot measures the sentiment of Rachel’s dialogue across three periods to track changes in her emotional expressions. Original sentiment scores include negative values, but after smoothing, the values are adjusted upwards, enabling comparisons across periods. Period one and two show greater variability, while the last period shows stability. On average, period three reflects more positive emotions than. Thus, Rachel’s emotional expressions in the scripts become progressively more positive and stable, mirroring improvements in her career and life as the narrative progresses.



Figure 13: Comparison Wordclouds between genders

The word cloud comparison reveals a clear gender disparity in character representation. Female characters frequently discuss aspects of their personal lives, with terms like "wife," "mom," "marriage," and "weddings" predominating. In contrast, male characters exhibit no distinct thematic preference, though their roles often appear more lighthearted, as indicated by recurring words like "duck" and "game," which are associated with humorous storylines in the show.

## 4 Friends as an English Language Learning Resource

### LSE candidate number: 24246

The sitcom "Friends" has long served as a valuable resource for non-native English speakers seeking to enhance their language skills. Divided into two subsections, the first conducts a statistical analysis to evaluate the language's diversity and complexity in the series, exploring metrics like Token-to-Type Ratios, Mean Length of Utterance, and Idioms Frequency. The second examines how language diversity and complexity vary across different thematic contexts within the show. This analysis includes episodes within the top and bottom 20% thresholds in terms of viewership and rating, aiming to explore the potential benefits and drawbacks experienced by "Friends" English learners.

### 4.1 The Diversity and Complexity of Language in Friends

In terms of viewership, Table 5 shows that both the most and least viewed episodes show an average TTR score of approximately 9.4%, indicating a similar high level of diverse vocabulary usage. However, the latter demonstrate a higher average MLU compared to the former, suggesting potentially more complex language structures.

In terms of ratings, Table 5 shows that although the average TTR score remains consistent for both the highest and lowest rated episodes, a notable difference emerges when examining the MLU. The former feature utterances that are, approximately, on average, one word longer than those in the highest ranked episodes. This suggests that, despite their lack of popularity among viewers, the least ranked episodes offer additional opportunities for English learners to encounter longer instances of spoken language.

The findings highlight the Friends sitcom's role as a rich resource for English language learners, with its TTR ratio approaching 1. Additionally, the average MLU of approximately 10 words per utterance aligns closely with the typical minimum spoken English utterance length for native English speakers (Nippold (2006)).

	<b>Most Viewed</b>	<b>Least Viewed</b>	<b>Highest Rated</b>	<b>Lowest Rated</b>
Average TTR	0.94	0.95	0.94	0.94
Average MLU	10.07	10.50	9.92	10.72

Table 5: Comparison of Viewership and Ratings

Next, we analyze the frequency of idioms used in the sitcom, utilizing a wordlist of static idioms sourced from the English Possible Idiomatic Expressions corpus (Saxena and Paul (2020)). Findings presented in Table 16 in Appendix reveal that the frequency of idioms varies across different categories of episodes. Specifically, there were 38 idioms found in the lowest rated episodes, 29 in the most viewed and highest rated, and 18 in the least viewed. Notably, idioms such as "big deal," "how come," and "you bet" exhibit significantly higher frequencies across all sampled episodes. This suggests that while learners may encounter a substantial number of idioms from watching the most viewed and highest rated episodes, they still have the opportunity to enhance their English vocabulary by watching the least viewed, and especially, lowest rated episodes.

## 4.2 Variations in Language Diversity and Complexity Across Thematic Contexts

### 4.2.1 Sentiment-Vocabulary analysis

The Friends dataset includes a sentiment classification for each utterance in the script. To explore potential variations in language diversity and complexity across different emotions, we compared TTR and MLU scores for each sentiment class. This analysis offers insights into how language varies across emotional contexts, which can be valuable for English learners as they develop appropriate vocabulary for different emotional settings.

Our findings, depicted in Tables 6-9, indicate that while the TTR scores exhibit minimal variation across the eight emotions analyzed, the MLU scores display more significant differences. This suggests that while the vocabulary diversity remains relatively consistent across emotional settings, the complexity of language structure varies notably. For instance, across all emotions, the MLU tends to be highest in settings associated with intense emotions such as sadness, anger, power, and fear. On the other hand, emotions like neutrality and peacefulness tend to have lower MLU scores. However, when it comes to TTR scores, the variation across emotions is less pronounced, indicating a consistent level of vocabulary diversity irrespective of emotional context. Visual representation of the patterns of vocabulary usage across sentiments are found in Figures 12-19 in the Appendix.

Table 6: Sentiment-Vocabulary Analysis of Most Viewed Episodes

<b>Emotion</b>	<b>Average TTR</b>	<b>Average MLU</b>
Joyful	0.91	11.76
Mad	0.92	13.65
Neutral	0.96	7.36
Peaceful	0.93	12.08
Powerful	0.93	12.36
Sad	0.93	14.42
Scared	0.91	11.68
NA	0.94	9.67

Table 7: Sentiment-Vocabulary Analysis of Least Viewed Episodes

<b>Emotion</b>	<b>Average TTR</b>	<b>Average MLU</b>
Joyful	0.93	10.5
Mad	0.93	12.33
Neutral	0.96	7.78
Peaceful	0.94	10.23
Powerful	0.91	12.95
Sad	0.94	13.94
Scared	0.92	9.41
NA	0.95	10.55

Table 8: Sentiment-Vocabulary Analysis of Lowest Rated Episodes

Emotion	Average TTR	Average MLU
Joyful	0.94	10.67
Mad	0.92	13.68
Neutral	0.96	7.91
Peaceful	0.94	11.73
Powerful	0.94	11.11
Sad	0.91	13.75
Scared	0.91	12.05
NA	0.95	10.74

Table 9: Sentiment-Vocabulary Analysis of Highest Rated Episodes

Emotion	Average TTR	Average MLU
Joyful	0.94	10.72
Mad	0.93	13.46
Neutral	0.96	7.61
Peaceful	0.94	11.03
Powerful	0.92	12.85
Sad	0.93	13.24
Scared	0.93	10.37
NA	0.95	9.84

#### 4.2.2 Context Dictionary Analysis

While the Friends sitcom serves as a potential resource for English language exposure among learners, it’s crucial to assess the effectiveness of this exposure across different linguistic contexts. We constructed a context-based dictionary by combining wordlists from the New General Service List Project (Browne, Culligan, and Phillips (2013)). This dictionary categorizes the 7868 most frequently used words in general, spoken, business, academic, and TOEIC English into labeled categories.

Our results, presented in Tables 10-13, reveal that across all sampled episodes, the sitcom employs approximately 50-54% of spoken and 41% of general high-frequency English words. This suggests that the show serves as an effective resource for acquiring vocabulary commonly used in everyday conversational English, aligning with its narrative premise.

Interestingly, TOEIC words are most prevalent in the least viewed and highest rated episodes, constituting 6% and 4% of the vocabulary respectively. Conversely, business-related terminology exhibits the lowest frequency in both the highest and lowest rated episodes, while being most frequent in the least viewed episodes, accounting for 1.7% and 2.2% respectively. Moreover, academic vocabulary is notably absent in the least viewed episodes, with the lowest frequency observed in the lowest rated episodes and the highest frequency in the most viewed episodes, comprising 0.5% and 2.7% respectively.

Table 10: Context dictionary of most viewed episodes

Context	Frequency
academic	2.7%
business	2.0%
general	41.7%
spoken	50.6%
toeic	2.7%

Table 11: Context dictionary of least viewed episodes

Context	Frequency
business	2.2%
general	41.4%
spoken	50.3%
toeic	5.9%

Table 12: Context dictionary of highest rated episodes

Context	Frequency
academic	1.7%
business	1.7%
general	41.1%
spoken	51.1%
toeic	4.1%

Table 13: Context dictionary of lowest rated episodes

Context	Frequency
academic	0.5%
business	1.7%
general	41.6%
spoken	53.7%
toeic	2.3%

## 5 Unveiling Criticism: Analysis of Offensive Content in Friends

### LSE candidate number: 24980

The show is often facing harsh criticism due to its inappropriate jokes about sensitive topics. This section of the analysis aims to answer the following question: How does the text reflect the criticism of the series, particularly regarding its homophobic, fat-shaming, or sexist content? This section consists of two main parts: first, an overall analysis of offensive language in Friends, and second, a focused analysis on fat shaming against Monica’s character.

As part of the pre-processing steps, numbers, punctuation and stopwords (including words such as “umm”, “ooh” etc.) have been removed and the words have been stemmed before creating the document-feature matrix (dfm). Some sub-questions required trimming the dfm to reduce features and weighting to obtain proportions.

### 5.1 General Analysis of Offensive Content

The overall analysis of offensive language is targeted to understand how homophobic, sexist, and fat-shaming content appears in the script. This section primarily uses self-defined dictionaries, “key words in context” and sentiment analysis. It is important to mention that there is no direct hate speech in the script, only references and comments, thus it is challenging to detect all of these using dictionaries.

Regarding homophobic content, through inspecting the key words (lesbian, gay etc.) in context one can conclude that the topic mostly appears as part of sarcastic jokes. Ross mentions this topic the most due to his ex-wife, but this topic is hardly the main focus even in episodes where it is mentioned frequently. The average sentiment score of sentences mentioning homosexuality is positive (1.53) and the most relevant homosexual character (Carol, ex-wife of Ross) is not mentioned in a negative context. Nevertheless, the fact that this topic is the subject of several jokes does reflect homophobic tendency.

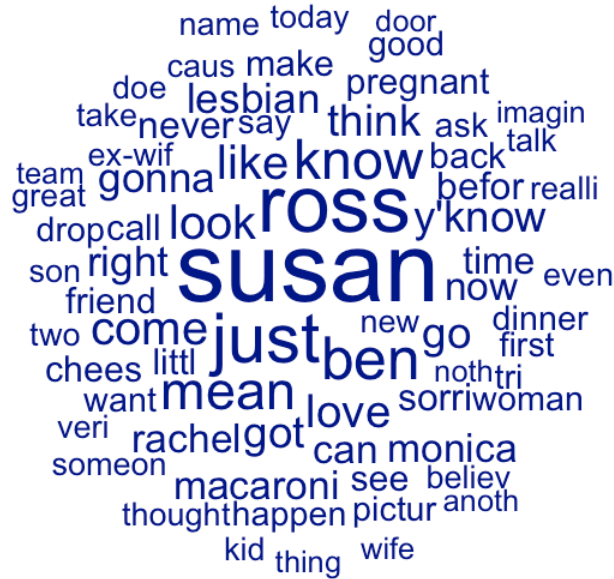


Figure 14: Most relevant words in the context of most relevant homosexual character, Carol (ex-wife of Ross)

Sexist references can be better observed in the script, as the word “female” appears in the context of sexuality, while “male” appears when describing non-traditional male careers (nanny, nurse etc.). Additionally, the objectification of women is reflected in the frequent mention of their looks and bodies. Using a self-defined dictionary one can observe that the male main characters use sexist language significantly more than the female main characters. Even though Joey’s character is criticized for objectifying women in the article of Wolf (2018), his scripts do not seem to reflect this tendency.

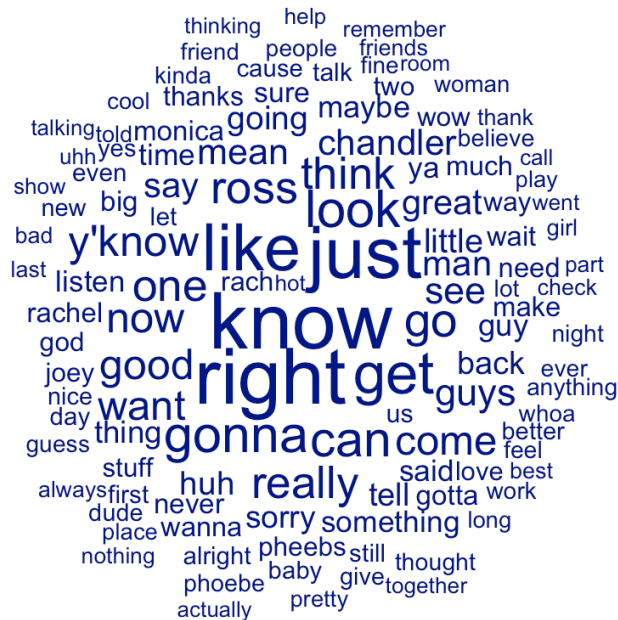


Figure 15: Despite criticism, Joey does not seem to mention objectifying content frequently

Insensitivity towards weight is mostly targeted against Monica due to her past issues, but other characters are also subjects of fat shaming. The mean sentiment score of texts mentioning the words “fat” or “weight” is slightly positive





## 6 Conclusion

This report revealed that *Friends*' lasting popularity stems from an interplay of distinctive characters, relatable themes, language learning potential, and its reflection of evolving social norms, despite criticisms of occasional offensive content.

While fine-tuned language models outperform traditional approaches more in conveying emotions than differentiating character speeches, their performance remains limited especially for character detection. The show's appeal lies not just in the distinctiveness of individual characters, but in the complex interplay between them that is challenging to capture through text alone. Another challenge is the characters' personalities and catchphrases changed over the course of 10 seasons. In contrary, emotions are more universal and less dependent on the specific context of the show. Emotions are often conveyed through specific words and phrases, which can be more easily detected by language models.

An exploration on the character development of Rachel shows some examples of the women characters representations in *Friends*. The core topic of the show, a story around a group of friends decided that the main theme was around the relationships between the characters. Rachel as the female representative left a positive impression through the multiple descriptions from the writers, no matter on scripts statistics, character relationships or sentiment analysis. She showed her determination into her career and paid consistent effort into it. One drawback from this analysis is that the three periods division is based on known storyline. And this storyline is based on her career-development. A more comprehensive analysis with personal-side of her life can be considered.

The third section underscores that while the show's excels in providing exposure to everyday conversational English, viewers of the least viewed episodes may encounter a rich vocabulary of TOEIC and business-related terms, whereas those of the most viewed episodes may encounter a broader array of academic terminology. This section focuses solely on textual dialogue, overlooking visual cues, scene directions, and translated subtitles into other languages, which could significantly affect language learning experiences. Additionally, the analysis does not consider the influence of popular culture topics and nouns on language acquisition, suggesting avenues for future research to explore the broader socio-cultural context of language learning through media consumption.

The last part of the analysis explored the offensive content of the show with a special focus on the reoccurring fat shaming against Monica's character. Although evidence supporting sexist, homophobic or fat shaming content has been found, these conclusions are limited by the applicability of dictionary methods in this case. As offensive content is most prevalent in forms of "creative" jokes, this method cannot detect sarcastic comments well. Further inaccuracies can occur as these topics are rare in absolute terms and not the most important traits for the concerned characters. These observations call for additional analysis with methods that take context into account, such as neural networks.

## References

- Browne, C., B. Culligan, and J. Phillips. 2013. “The New General Service List.” <http://www.newgeneralservicelist.org>.
- Cancella, Maju. 2020. “Friends: Every time a character was fat-shamed (ranked least to most offensive).” Accessed April 24, 2024. <https://screenrant.com/friends-characters-fat-shamed/>.
- Cobb, S., N. Ewen, and H. Hamad. 2018. “Friends Reconsidered: Cultural Politics, Intergenerationality, and Afterlives.” *Television New Media* 19 (8): 683–691. Accessed April 26, 2024. <https://doi.org/10.1177/1527476418778426>.
- EmilHvitfeldt. n.d. “EmilHvitfeldt/R-text-data: List of textual data sources to be used for text mining in R.” Accessed April 25, 2024. <https://github.com/EmilHvitfeldt/R-text-data?tab=readme-ov-file#friends>.
- Holbert, R. Lance, Dhavan V. Shah, and Nojin Kwak. 2003. “Political Implications of Prime-Time Drama and Sitcom Use: Genres of Representation and Opinions Concerning Women’s Rights.” *Journal of Communication* 53 (1): 45–60. Accessed April 26, 2024. <https://doi.org/10.1111/j.1460-2466.2003.tb03004.x>.
- Larrea Espinar, Antonio R., Angela María Raigón Rodríguez. 2020. “Sitcoms as a tool for cultural learning in the efl classroom.” *Revista de Medios y Educación*, 33–50. <https://idus.us.es/handle/11441/93731>.
- Nippold, M.A. 2006. “Language Development in School-Age Children, Adolescents, and Adults.” In *Encyclopedia of Language & Linguistics*, edited by Keith Brown, 368–373. Elsevier. ISBN: 9780080448541. <https://doi.org/10.1016/B0-08-044854-2/00852-X>. <https://www.sciencedirect.com/science/article/pii/B008044854200852X>.
- Saxena, Prateek, and Soma Paul. 2020. *EPIE Dataset: A Corpus For Possible Idiomatic Expressions*. arXiv: 2006.09479 [cs.CL].
- Tally, M. 2016. *The Rise of the Anti-Heroine in TV’s Third Golden Age*. Cambridge Scholars Publishing. Accessed April 26, 2024. <https://books.google.co.uk/books?hl=zh-CN&lr=&id=HuW2DQAAQBAJ>.
- Wolf, Taylor. 2018. “Joey from ‘Friends’ Lacks Respect for Women - The Baylor Lariat.” Accessed April 24, 2024. <https://baylorldariat.com/2018/09/13/joey-from-friends-lacks-respect-for-women/>.
- Zeisler, Andi. 2016. *We Were Feminists Once: From Riot Grrrl to CoverGirl®, the Buying and Selling of a Political Movement*. PublicAffairs. Accessed April 26, 2024. <https://books.google.co.uk/books?hl=zh-CN&lr=&id=nj1WDgAAQBAJ>.

## Appendix

Name	Introduction
Barry	Rachel's biance, she escaped from the wedding with him as she realised he is not the one she loved at the start of season 1. This movement marked her change from daddy's daughter as she have to live independently by herself in NYC.
Joshua	Joshua is her client when Rachel worked in the Bloomingdale's as personal shopper. Rachel and Joshua dated with each other before but broke up at the end.
Joanna	Joanna is Rachel's boss in Bloomingdale's as assistant shopper. She provided Rachel the job and she was one of the interviewers when Rachel applied for Bloomingdales's.
Zelner	Zelner is Rachel's boss in Ralph Lauren.

Table 14: Characters Introduction

Class Name	Words Contain
Work Related	"job", "work", "success", "congrats", "honor", "dream", "grow", "career", "interview"
Money Related	"rich", "salary", "money", "poor", "dollar", "wealth"
Career Related	"merchandise", "waitress", "fashion", "retail", "assistant", "executive", "director", "ralph", "laurent", "buyer"

Table 15: Self-defined Dictionary Method In Section Two

Idiom	Most Viewed	Least Viewed	Highest Rated	Lowest Rated
all set	0	0	1	2
all the way	7	9	8	16
alphabet soup	0	1	0	1
about face	0	0	0	1
at hand	0	0	0	1
a steal	0	0	0	1
big bucks	1	2	1	0
big deal	10	27	14	44
close call	0	2	0	0
cutting edge	0	0	0	1
dead meat	0	0	0	1
dead right	0	1	0	1
dead serious	1	0	1	0
fair enough	0	0	1	0
fall for it	0	0	1	0
filthy rich	0	0	0	1
flea market	0	0	0	7
give me five	1	1	1	1
god forbid	3	0	1	0
god knows	0	0	0	1
green light	1	0	1	0
gut feeling	1	0	1	0
hold on a second	5	4	6	3
how come	16	5	14	27
in the bag	2	2	2	2
in the zone	1	0	3	0
just in case	1	3	1	2
last but not least	0	0	0	2
never mind	0	1	1	5
no rush	0	0	0	1
no strings attached	0	0	0	1
now and then	2	3	1	3
off the hook	0	0	1	0
on the rocks	2	2	2	1
on the same page	0	0	0	1
on the table	2	2	1	7
once in a blue moon	0	0	0	1
out of sight	0	0	0	1
out of the blue	1	0	0	0
per se	2	0	1	2
piece of cake	0	0	0	2
pull yourself together	1	0	1	0
quality time	2	0	0	1
rough time	1	0	0	0
round the clock	0	0	0	1
silver lining	2	0	0	0
son of a gun	1	0	1	0
sooner or later	2	0	0	2
thank goodness	0	0	0	1
thumbs up	1	0	1	0
time is money	0	0	0	2
washed up	0	0	1	0
water under the bridge	0	1	0	0
way around	3	2	0	0
well read	1	0	0	0
you bet	10	3	11	13
zero hour	0	0	1	0
<b>Total</b>	29	18	29	38

Table 16: Combined Table of Idioms and Their Frequencies

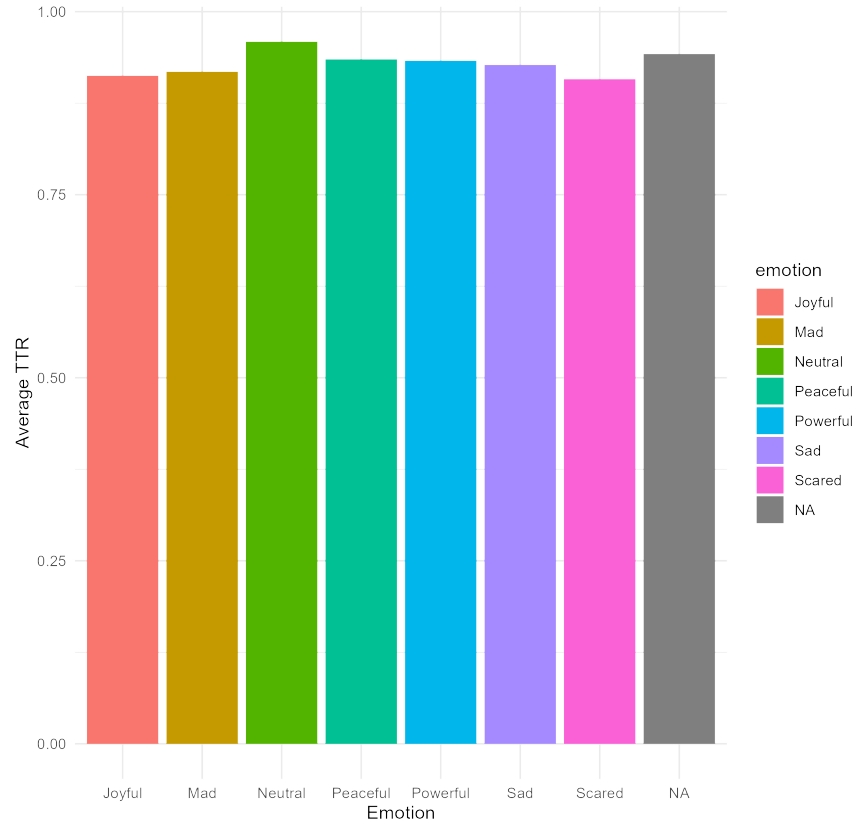


Figure 17: Average TTR for Most Viewed Episodes

Character	Prevalence of topic of homosexuality
Chandler Bing	0.0003505082
Joey Tribbiani	0.0001528585
Monica Geller	0
Phoebe Buffay	0.0010498688
Rachel Green	0.0005767844
Ross Geller	0.0010794140

Table 17: Mentions of the Topic of Homosexuality among Friends Main Characters

	feature <chr>	chi2 <dbl>	p <dbl>	n_target <dbl>	n_reference <dbl>
1	think	307.7244983	0.000000000	8.8333333	4.3967014
2	mistak	10.6267525	0.001114633	0.3333333	0.2686869
3	shoot	6.4352059	0.011188019	0.3333333	0.6250000
4	dear	3.7317366	0.053387859	0.5000000	2.7651515
5	star	2.9970211	0.083417765	0.2500000	0.8472222
6	woman	1.7798995	0.182161665	0.2500000	1.4436508
7	actual	1.6405469	0.200250453	0.3333333	2.6690657
8	fat	1.4685607	0.225572920	0.2500000	1.7192460
9	guy	1.4207736	0.233276119	0.6666667	9.8435961
10	thousand	1.4008975	0.236573355	0.2500000	1.7916667

Figure 25: Cluster Mentioning the Topic of Weight in K-means Clustering

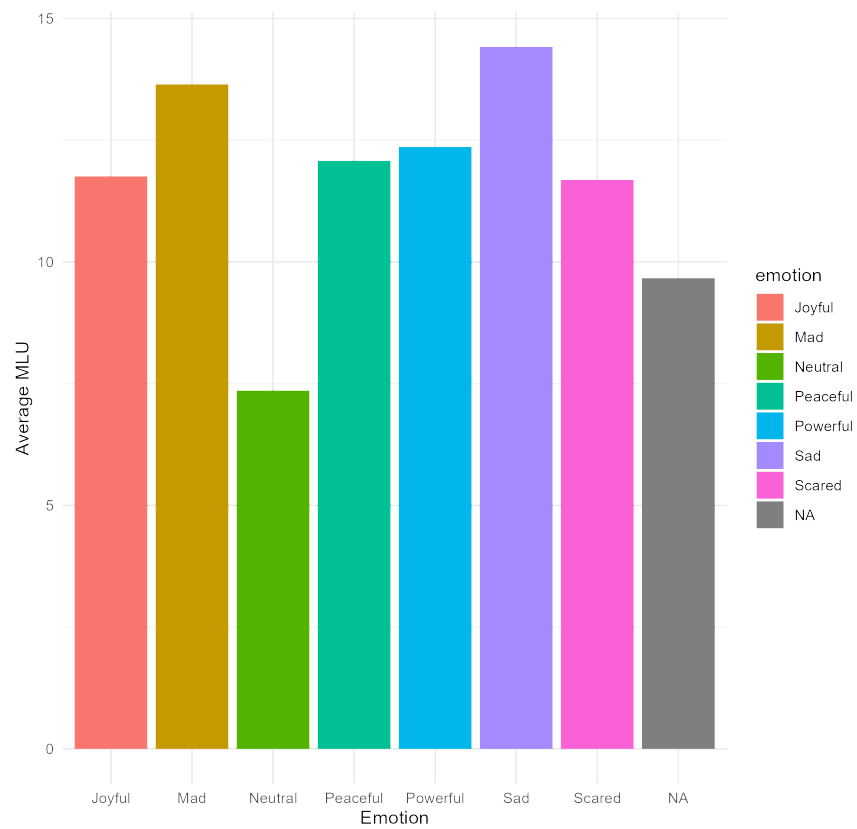


Figure 18: Average MLU for Most Viewed Episodes

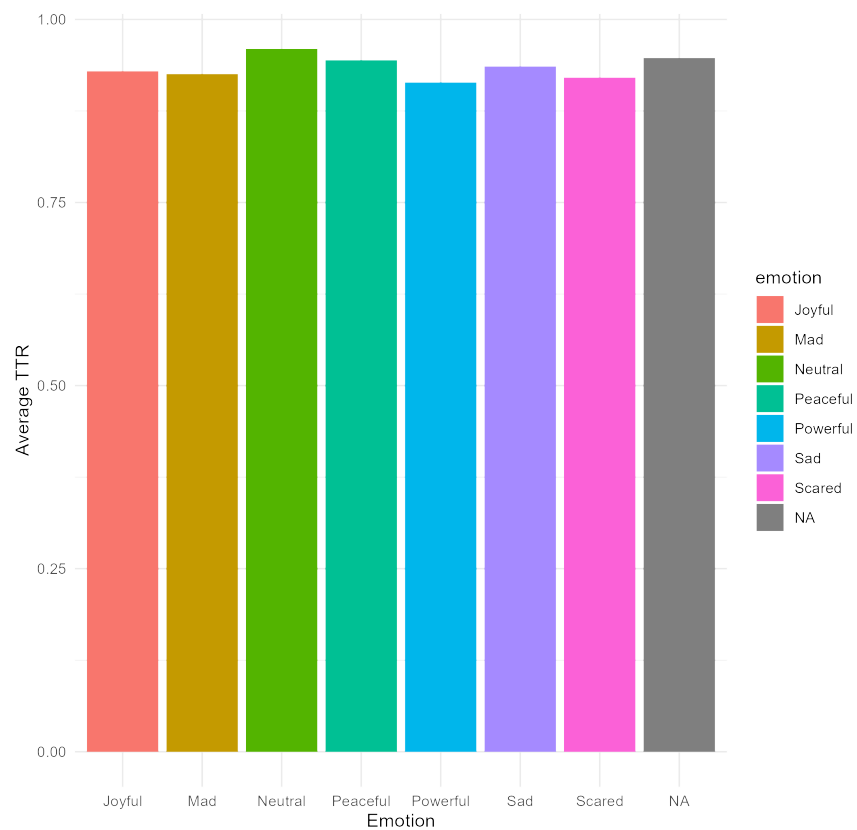


Figure 19: Average TTR for Least Viewed Episodes

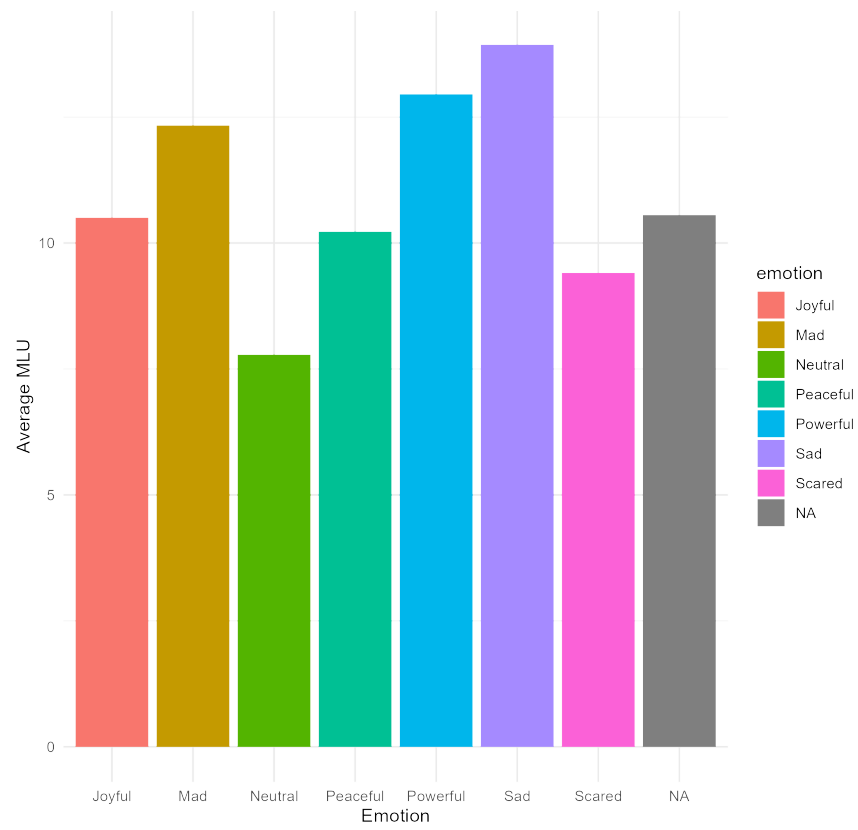


Figure 20: Average MLU for Least Viewed Episodes

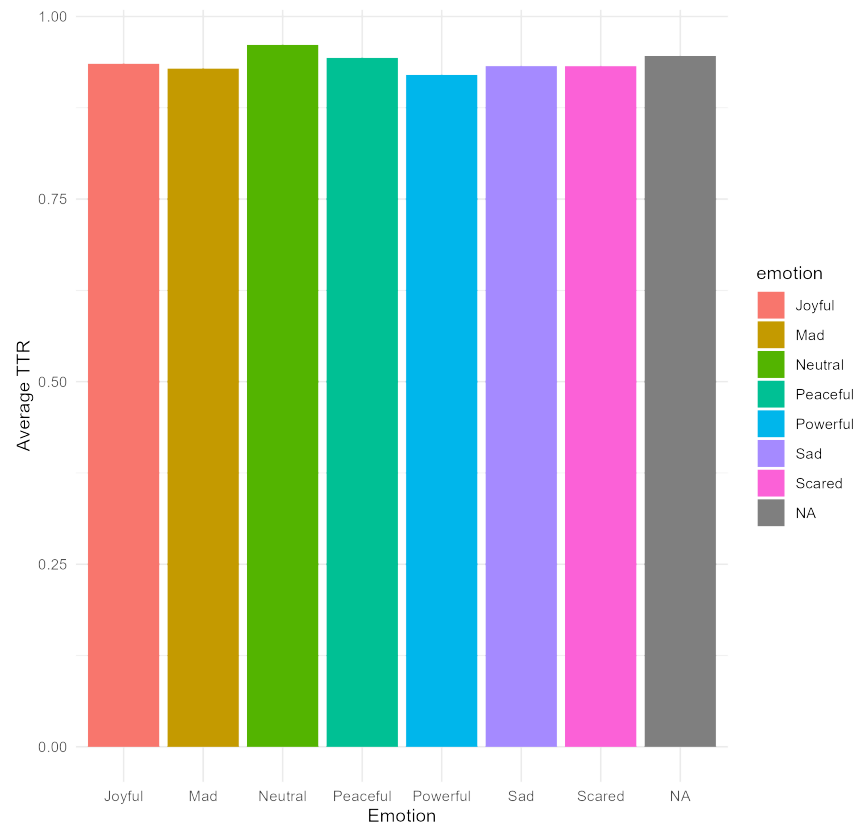


Figure 21: Average TTR for Highest Rated Episodes



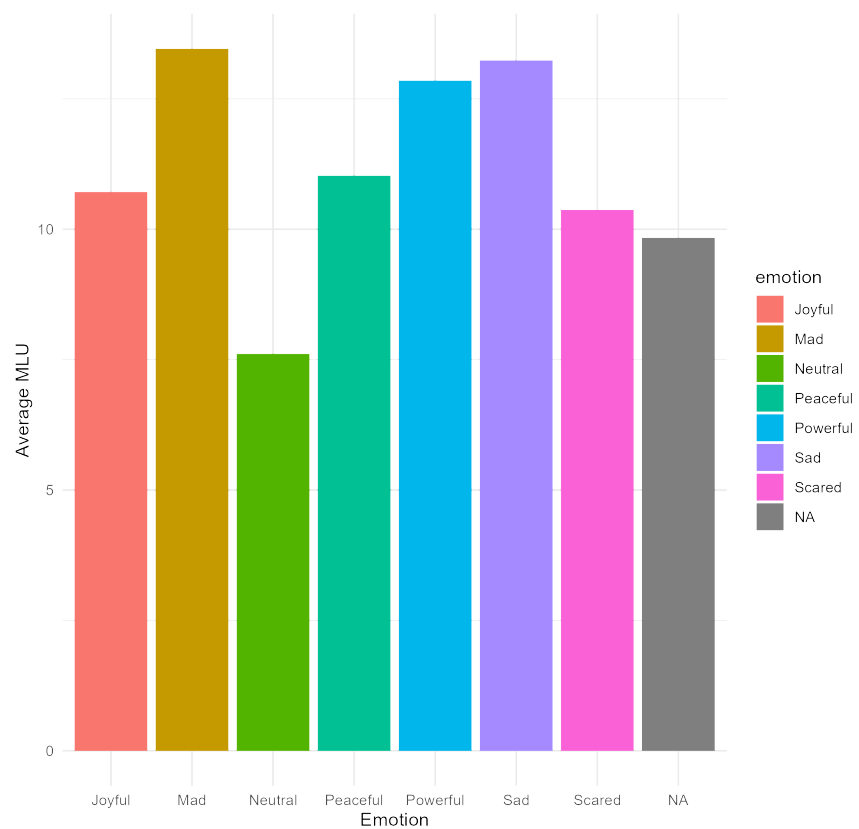


Figure 22: Average MLU for Highest Rated Episodes

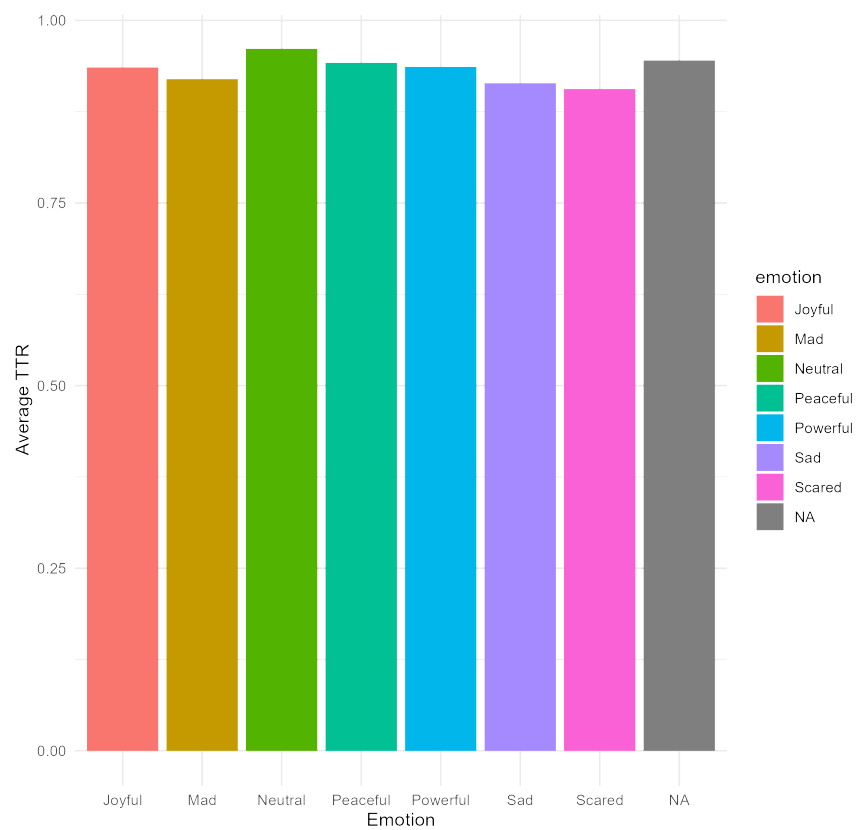


Figure 23: Average TTR for Lowest Rated Episodes

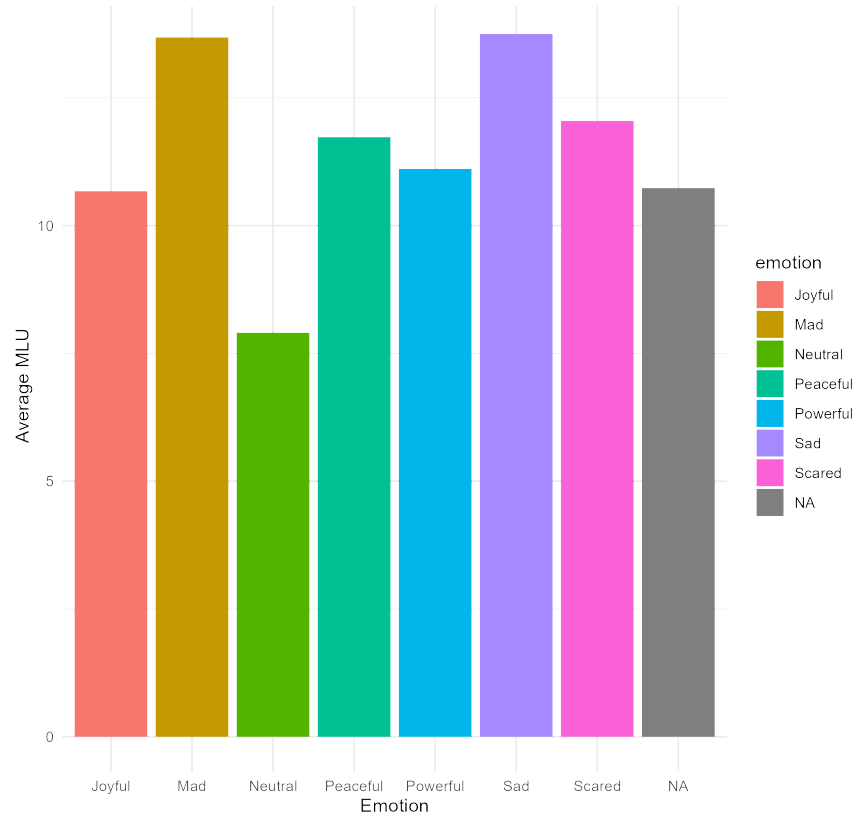


Figure 24: Average MLU for Lowest Rated Episodes

Episode Name	Prevalence of topic of homosexuality
The One with Phoebe's Husband	0.005982054
The One with the Flashback	0.005279831
The One with the Memorial Service	0.004930156
The One with Rachel's Assistant	0.004642526
The One Where Paul's the Man	0.003925417
The One with the Birth	0.003917728
The One with Joey's Interview	0.003864734
The One Where Nana Dies Twice	0.003726708
The One with Monica's Thunder	0.003679853
The One with Rachel's New Dress	0.003663004

Table 18: Top 10 episodes Mentioning the Topic of Homosexuality the most

Character	Sexist language
Chandler Bing	0.0028040659
Joey Tribbiani	0.0024457352
Monica Geller	0.0005172414
Phoebe Buffay	0.0006999125
Rachel Green	0.0012977650
Ross Geller	0.0023130301

Table 19: Prevalence of Sexist language of Main Friends Characters

Episode Name	Sexist language
The One with the Candy Hearts	0.013253012
The One with the Engagement Picture	0.010357815
The One with the Breast Milk	0.009557945
The One with Phoebe's Husband	0.008973081
The One Where Monica Sings	0.008708273
The One with the Flashback	0.008447730
The One with the Butt	0.008064516
The One Where No One Proposes	0.007936508
The One with the Sharks	0.007858546
The One with the Halloween Party	0.007662835

Table 20: Top 10 episodes using Sexist Language the most

Character	Fat shaming Language
Chandler Bing	0.0007010165
Joey Tribbiani	0.0004585754
Monica Geller	0.0003448276
Phoebe Buffay	0.0003499563
Rachel Green	0
Ross Geller	0.0004626060

Table 21: Prevalence of Sexist language of Main Friends Characters

Episode Name	Fat shaming Language
The One Where No One's Ready	0.007544007
The One with the Nap Partners	0.005747126
The One with All the Thanksgivings	0.005736138
The One with the Jam	0.003875969
The One Where Ross Finds Out	0.003092784
The One with the List	0.002116402
The One with the Giant Poking Device	0.001986097
The One with the Chicken Pox	0.001980198
The One That Could Have Been	0.001937046
The One Where the Monkey Gets Away	0.001906578

Table 22: Top 10 episodes using Fat Shaming Language the most

## Top Topics

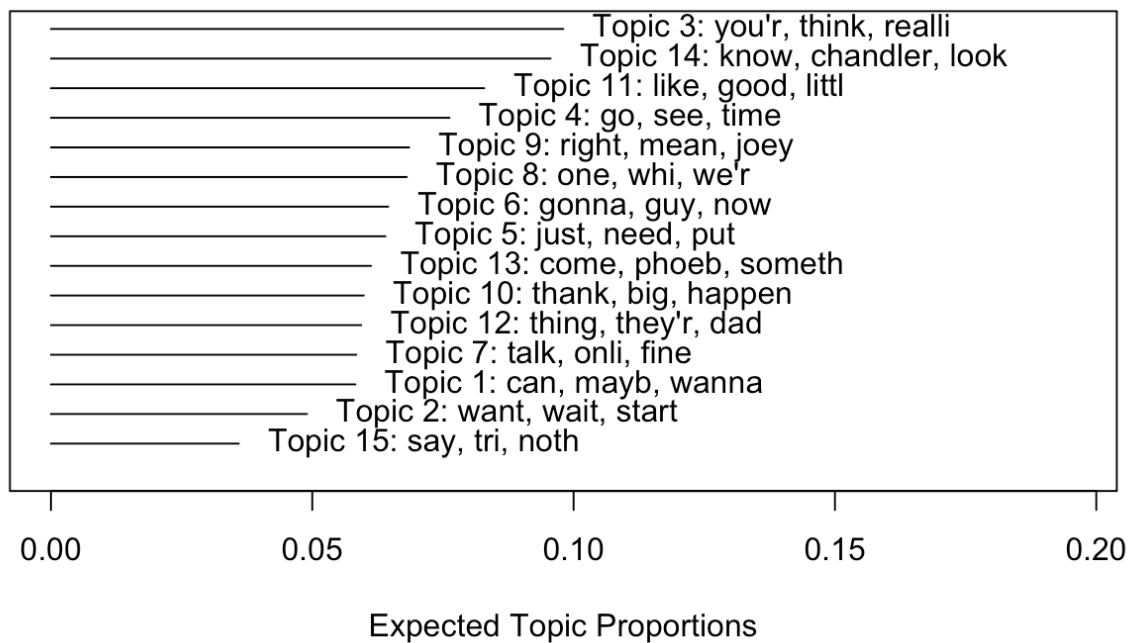


Figure 26: Top 15 Topics Monica Mentions

## Top Topics

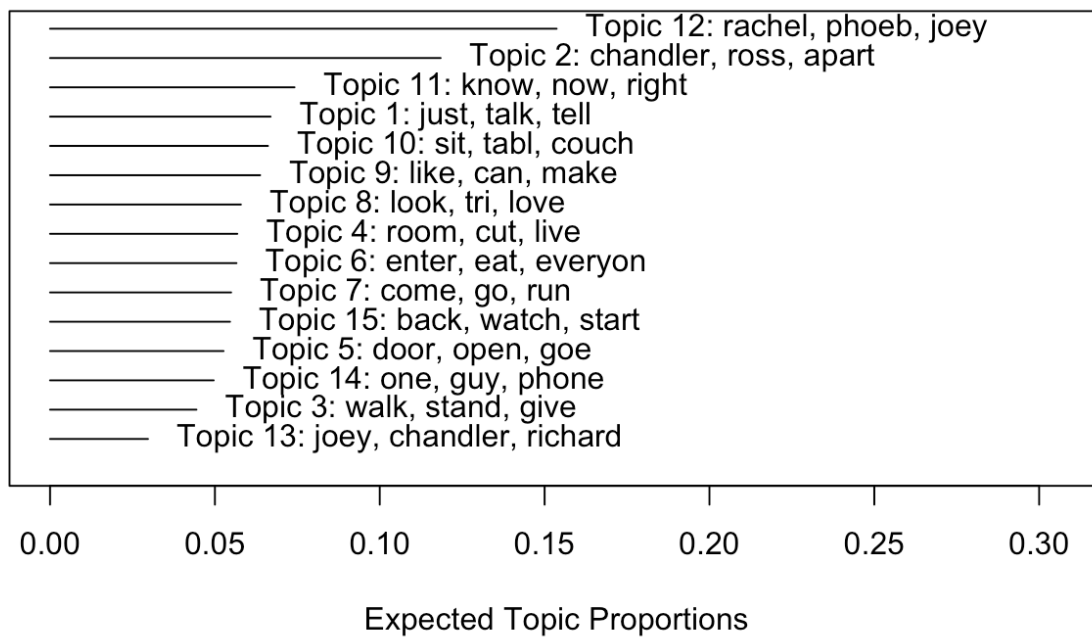


Figure 27: Top 15 Topics Other Characters Mention about Monica