

## **Naive Text Analysis of 'The Adventure of the Six Napoleons'**

*Text/sentiment analysis of A. C. Doyle's short story featuring the popular literary character, Sherlock Holmes.*

*Karn Saheb — [karn.saheb@mail.utoronto.ca](mailto:karn.saheb@mail.utoronto.ca)*

Illustration c/o Mariana F. — [dribbble.com/marina-f](https://dribbble.com/marina-f)

## Preface: “It’s elementary, my dear Watson!”

### *The Adventure of the Six Napoleons*

- A. C. Doyle tells a short story of a curious case of a serial burglar turned murderer and a number of broken replica busts (statues) of Napoleon the First.
- Seems like dialogue serves as the primary means of storytelling and narration only to supplement the conversations of the characters. This lend a unique voice to the characters, but it also develops their personas and serves to advance the plot dynamically<sup>1</sup>.
- Key insights:
  - Large imbalance of dialogue and narration, this adds more weight to the sentiments and how they communicate the larger narrative.
  - The sentiments (e.g. joy, fear, trust) of the characters in their interactions with Holmes.
- Casual conclusions about how certain attitudes expressed by the characters reinforce the anticipatory and suspenseful atmosphere of Doyle's short crime/mystery story.
- This is all preliminary findings, much more analysis and research of related works remains.

<sup>1</sup>“When characters start talking, the story starts moving” (Gloria Kempton, Dialogue 12).

## Research Question

What are the main sentiments of the various characters in A. C. Doyle's 'Sherlock Holmes: The Adventure of The Six Napoleons' and how do they relate to the larger narrative of the story?

## What Makes this Question Important?

Researching the sentiments identifies how the characters are perceived in a quantifiable manner, these sentiments shape the outward personalities of these characters<sup>2</sup>.

Identifying sentiment and personality questions can be important in areas such as author attribution and similarity to other works (fan-fiction, spinoffs, imitations of original works).

## Further Questions

- How do the sentiments of recurring characters change over the various Sherlock adventures and are they consistent?
- What does this consistency, if any, say about Doyle's methods of story telling?
- Can this process be applied to other such texts?

<sup>2</sup>“Personality detection from text means to extract the behavior characteristics of authors written the text” (Personality Detection from Text: A Review — Basant Agarwal 4).

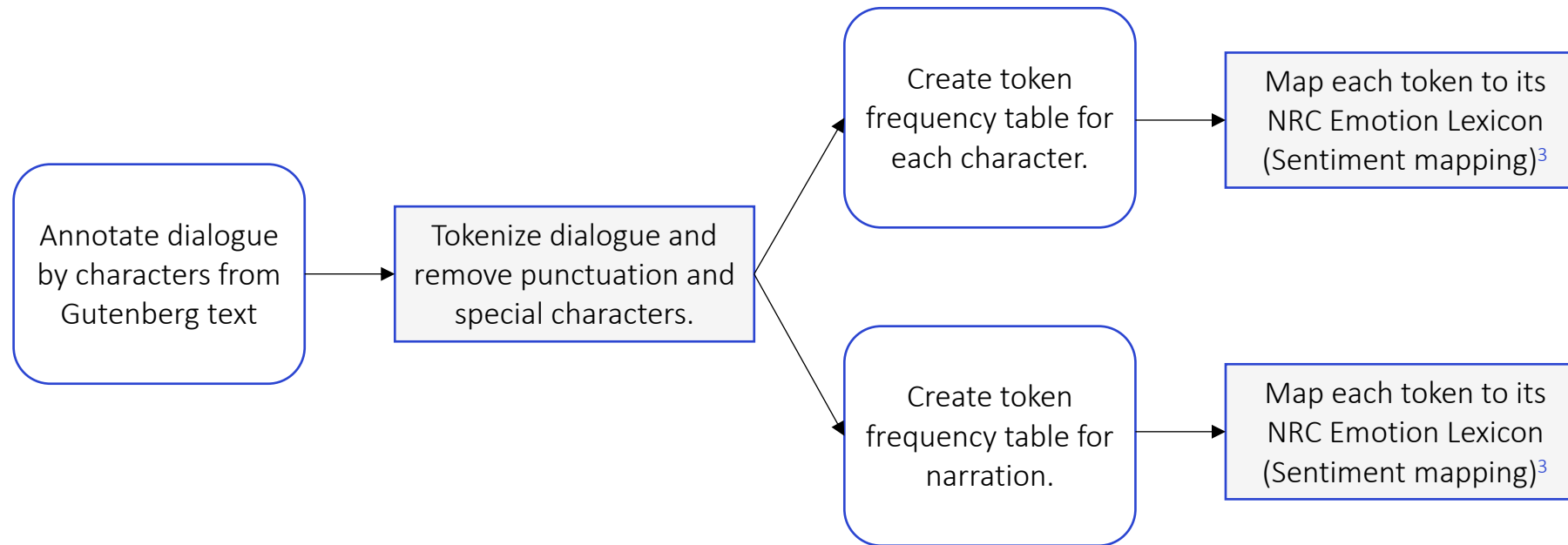
## Research Overview

A **naive** exploration of the sentiments of the various characters in short story, 'The Adventure of the Six Napoleons' through rudimentary text analysis. In addition, a short comparison of the proportion of conversation and dialogue within the story. Finally, a brief summary with some limitations and casual conclusions about the analysis.

The statistical programming language **R** was used to perform all the text processing, data analysis, and visualization. The source code will be made available for viewing (and criticism) in the next few days after it has been optimized and documented.



# Process



<sup>3</sup>There are a number of unigrams that appear in the text that do not appear in the NRC sentiment lexicon, thus there may be missing sentiment data.

## Data Overview

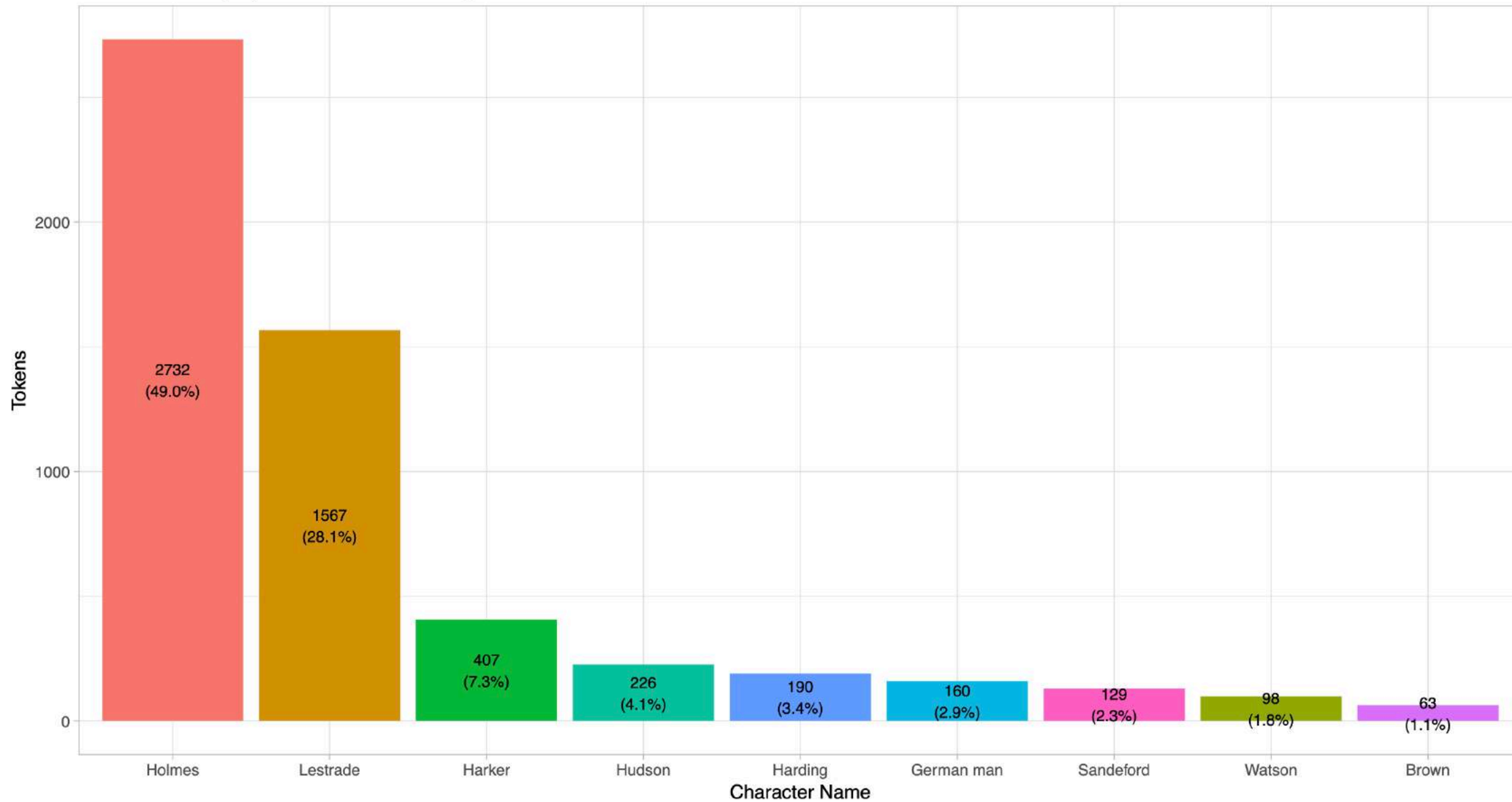
The resulting frequency tables were then visualized as follows:

- **Conversational Contributions by Character**: The number and proportion of tokens in each characters dialogue.
- **Conversational Sentiments by Character**: The proportion of the dialogue which was identified as a certain sentiment.
- **Average Dialogue Sentiment**: An average proportion of a given sentiment in all the character dialogue.
- **Proportion of tokens in Narration vs. Character Speech**: An direct comparison of tokens associated with dialogue and tokens associated with conversation.



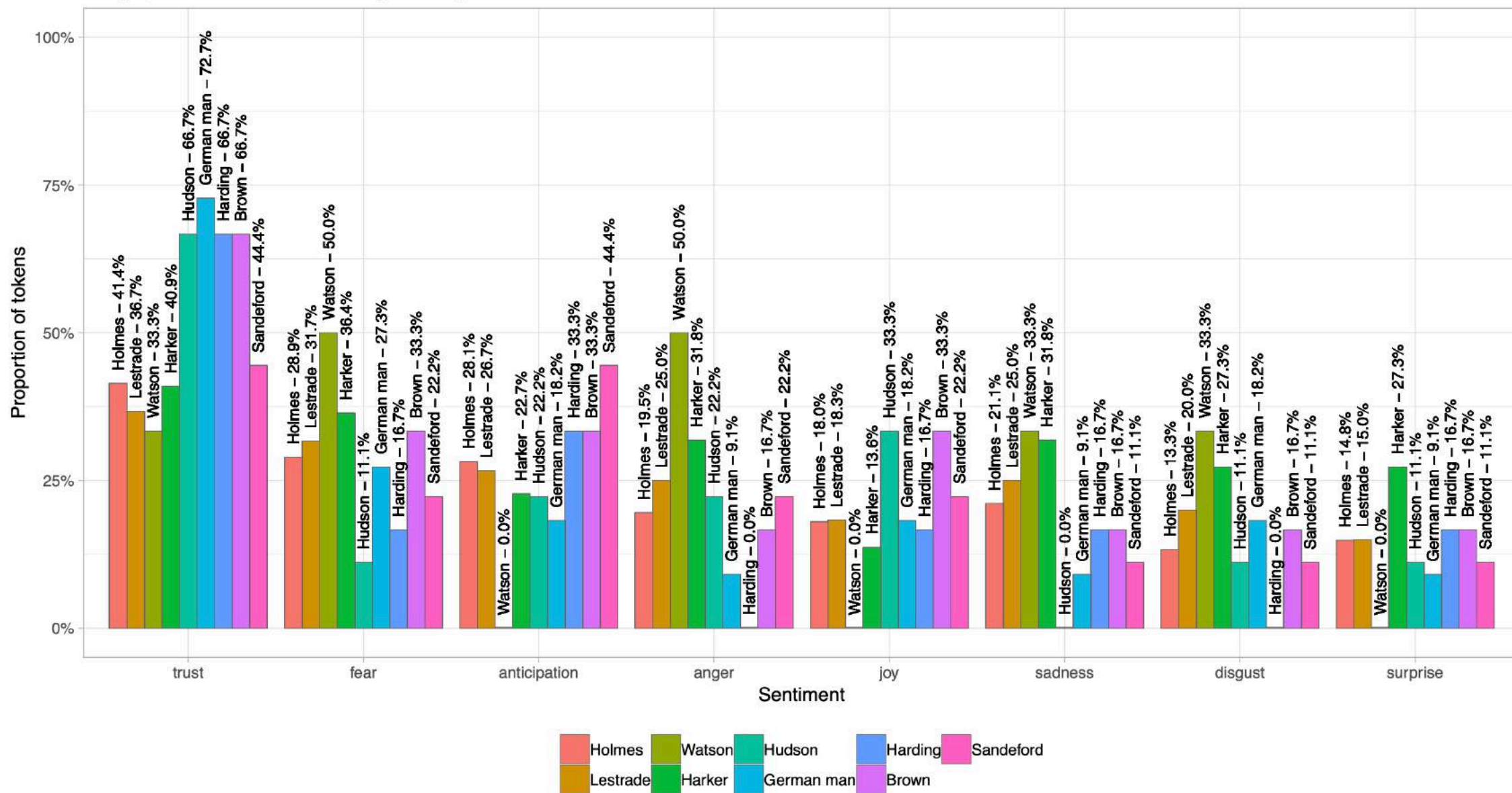
## Conversational Contribution by Character

The number and proportion of tokens in dialogue for each character.



## Conversational Sentiments by Character

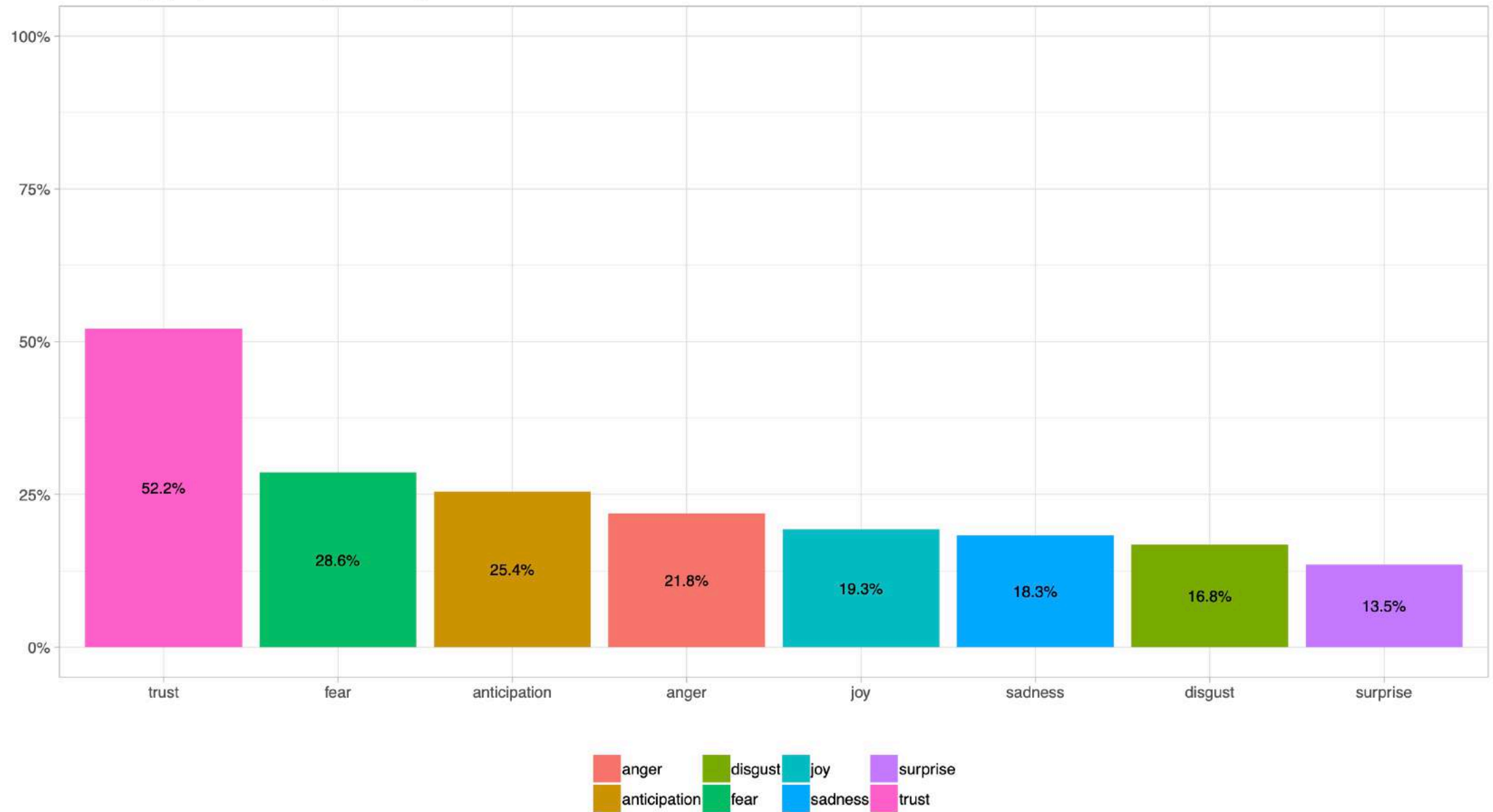
The proportion of character's dialogue with a given sentiment.





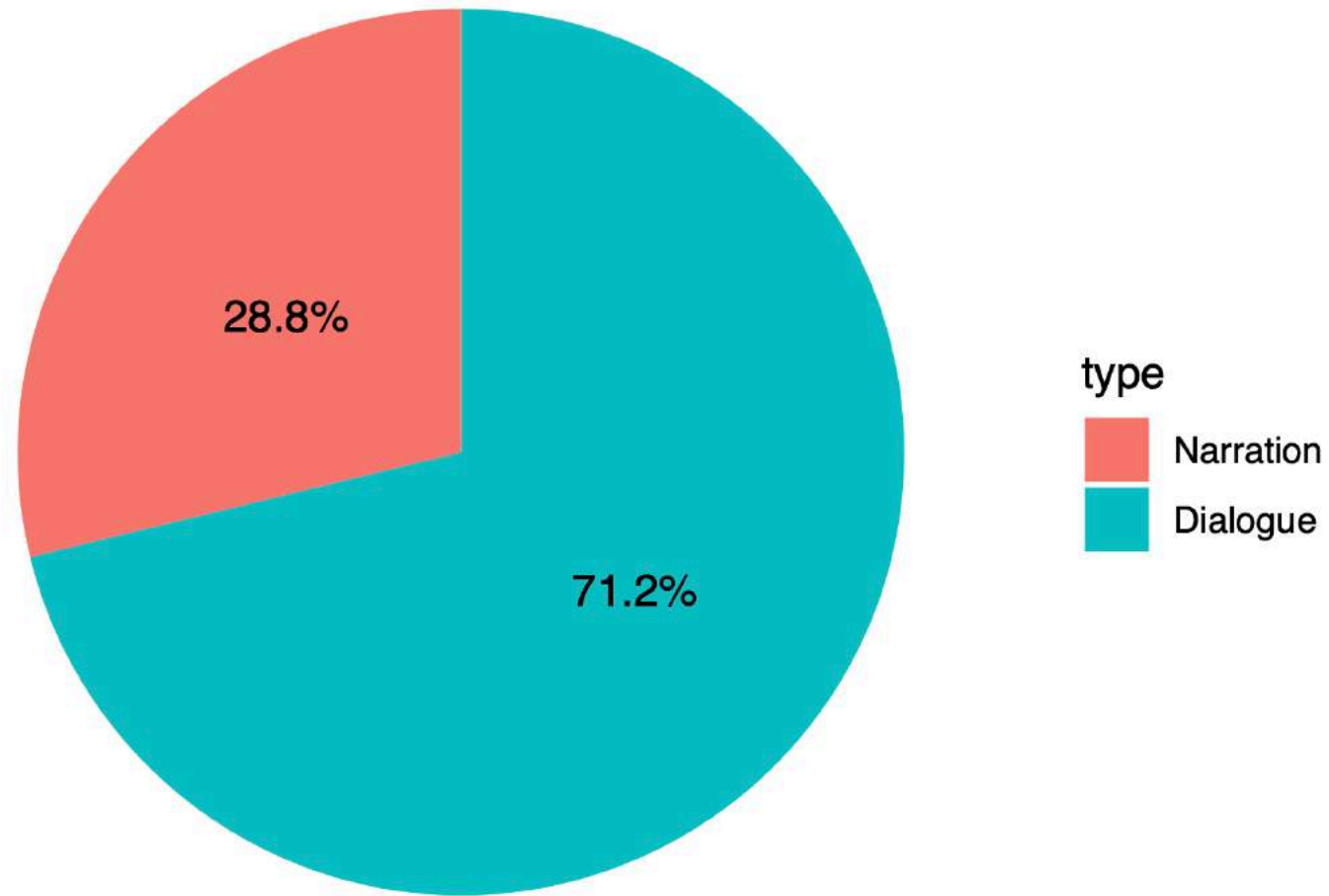
## Average Dialogue Sentiment

The average proportion of dialogue with a given sentiment.



## Proportion of Tokens in Narration vs. Character Speech

Comparing the proportion of tokens in speech.



## Conclusions

- Holmes and Lestrade make up the bulk of the conversation and Watson has almost no actual conversational contribution. Most of the characters have high levels of **trust** and **fear**, both of which can be attributed to the reverence of the character of Sherlock<sup>4</sup>, and the ongoing issue of a murderer on the loose.
- The **anticipation**, **fear** and **trust**, which are the three most prominent sentiments expressed by the characters are all telltale emotions which are pillars of a story within the genre of mystery. These sentiments help reinforce the themes of the genre, in particular the **anticipation** and **fear** are both used to attract the reader and build up the mystery.
- The visualization of the proportion of Dialogue and Narration indicate that Dialogue is the primary means of storytelling (**71%**) and narration (**29%**) is used only to supplement the conversations of the characters. This means that the sentiments of the characters make up a larger proportion of the story and as a result, more weight can be placed on the prominent sentiments as representation of the the larger story.

<sup>4</sup>“if you come down to-morrow, there’s not a man, from the oldest inspector to the youngest constable, who wouldn’t be glad to shake you by the hand” (Doyle).

## Limitations and Cautions

- This is a **small sample size** — the analysis only contains the data from a single story and has no baseline to measure the results of the text analysis. An application of the previously defined *Process* on other Holmes adventures would provide additional comparison data.
- The NRC Emotion Lexicon is based on a sentiment classification from the 2000s, for the purposes of a naive exploration of sentiments it is adequate but it **lacks the accuracy and breadth** that a lexicon that was from the same time period as the text would have had.
- The mapping of **sentiments are not aware of context** — some words such as *affection* may be considered to be associated with *joy* but when in a sentence like '*lack of affection*', the words means something totally different.
- Some tokens have no, or have an ambiguous sentiment associated with them and may **remove certain words from consideration** due to this lack of association.

## Future Work

- Explore other Holmes adventures to **determine trends and validate conclusions** about author style with respect to the narration and dialogue imbalance.
- Identify and implement a **more accurate sentiment mapping** for the tokens.
- Investigate **machine learning** implementations of the resulting classification of personality by distribution of sentiments to **identify if a particular character is speaking in a given conversation**.



**Thank You**

*[karn.saheb@mail.utoronto.ca](mailto:karn.saheb@mail.utoronto.ca), [hello@karn.io](mailto:hello@karn.io)*

*<https://karn.io>*

*<http://github.com/karn>*