

Variation in sentence length between novels by Jane Austin and Arthur Conan Doyle

Daphne Meekers

s5611989

Abstract

This study examines the average sentence length in novels by Jane Austen and Arthur Conan Doyle, focusing on how their writing reflects their storytelling approaches. By comparing novels of these two authors, we aim to better understand if sentence length reflects readability, audience engagement, and the overall tone of a narrative. We hypothesize that Austen's novels use longer, more descriptive sentences that reflect her focus on social interactions and character development, while Doyle's novels will contain shorter, action-driven sentences to maintain a fast-paced narrative.

1 Introduction

Sentence length is a meaningful linguistic feature, it helps understanding how writing styles evolve over time and reflects cultural and literary norms. It provides insight into how sentence length affects different audiences, impacting readability, the reader's perception, and their appreciation for the literary works. These findings may assist in better understanding and appreciating these authors' styles.

Previous research has demonstrated the importance of sentence length in stylistic analysis. For example in a study by Muchnik (2002) sentence length is proved to be an indicator of narrative style and pace, by examining sentence length in two novellas. Sentence length can differ significantly between works, which influences how readers experience and interpret the written text. In the research of Matthews and Folivi (2023) is determined that shorter sentences improve readability and provide greater accessibility and inclusion,

which is also seen in fields such as public health communication.

This study aims to explore the research question: Does sentence length reflect the different narrative and storytelling styles of Jane Austen and Arthur Conan Doyle? The hypothesis is that Jane Austen's novels will contain longer sentences compared to Arthur Conan Doyle's novels. This will reflect the more elaborate and descriptive writing style of Austen, containing descriptive characters' thoughts, development, and social dynamics. To convey these details she uses longer, more elaborate and more complex descriptions, which will increase the length of the sentences. Doyle's style is designed for fast-readable detective stories, consisting of shorter, straightforward sentences that keep the action and the plot clear. Therefore, Austen's elaborate style will lead to longer sentences, while Doyle's action-focused writing is more concise and to the point.

2 Related Work

Previous research has extensively studied sentence length as a stylistic and functional feature. Muchnik (2002) investigated how sentence length impacts narrative style and pacing. The study demonstrates the relevance of sentence length in understanding and distinguishing literary styles. There has been found that longer sentences contribute to detailed and elaborate storytelling, while shorter sentences drive action-oriented narratives. This is a useful insight regarding the comparison of Austen's descriptive storytelling with Doyle's fast-paced detective fiction. Matthews and Folivi (2023) conducted a study on sentence length perception, according to the study shorter sentences improve readability and inclusivity. Their findings show that sentence length impacts accessibility, making texts more inclusive for diverse read-

ers. This aligns with Doyle’s accessible, action-oriented style and can be helpful in comparing the readability. These findings help to understand the importance of analyzing how sentence length influences narrative tone, readers engagement, and accessibility in the novels of Austen and Doyle. Further, research by Rudnicka (2018) revealed a broader perspective on sentence length. It shows that sentence length has decreased over time over different genres. This change is connected to shifts in society. For example, with more and more people being able to read, writers adapted their writing style to appeal wider audiences. Changes in punctuation over time also influenced writing styles. Rudnicka (2018) explains that sentence length varies widely between genres, reflecting differences in storytelling styles. For example, shorter sentences are often connected to simpler sentence structures. In contrast, longer sentences use more detailed and complex structures, enabling more focus on character development and social interactions. The findings by Rudnicka (2018) help in differentiating between short and long sentences in novels. Deveci (2019) supports the work by Rudnicka (2018), by identifying that sentences containing around 20 words are optimal for comprehension. While this study focused on sentence length in research and educational texts, the threshold will still be useful for analyzing sentence length in novels. We can apply this threshold to novels because understanding any type of text depends on how well readers can process the sentences and comprehend them.

3 Data

The dataset includes two novels from each of the authors investigated, Jane Austen and Arthur Conan Doyle. These novels are retrieved from Project Gutenberg, an online archive of public domain texts.

- Jane Austen: Sense and Sensibility (1811) Pride and Prejudice (1813)
- Arthur Conan Doyle: The Adventures of Sherlock Holmes (1892) The Hound of the Baskervilles (1902)

There are two variables included in the research:

- The dependent variable is the average number of words per sentence, categorized by short and long sentences.

- The independent Variable is the Author: Jane Austen or Arthur Conan Doyle.

The size of the dataset is the first 1000 words of each book. Python will be used for text processing, the extraction of sentences and words, and the counting of the words. The threshold for differentiating short from long sentences will be as follows: short sentences contain fewer than 20 words, and long sentences contain 20 words or more. Table 1 visualizes how the data will be stored and viewed in this study.

Author	Novel	Short Sentences (< 20 words)	Long Sentences \geq 20 words
Jane Austin	Sense and Sensibility (1811)		
Jane Austin	Pride and Prejudice (1813)		
Arthur Conan Doyle	The Adventures of Sherlock Holmes (1892)		
Arthur Conan Doyle	The Hound of the Baskervilles (1902)		

Table 1: Visualization of the data storing structure

Chosen Threshold The threshold of 20 words was chosen based on trends observed in historical sentence lengths, as discussed by Rudnicka (2018) and was based on Deveci (2019), who studied how sentence length affects how we process text. This threshold of 20 words per sentence reflects the idea that we can only handle so much information at once, whereas longer sentences require more mental effort. While the research of Deveci (2019) is focused on educational and research texts, the same principle applies to other types of writing, including novels. The analysis of sentence lengths in fiction genres over time, particularly in the 19th and 20th centuries, indicates that sentences averaging 20 words or more are associated with more elaborate and descriptive writing. By contrast, shorter sentences often

enhance the readability and action-oriented narratives (Rudnicka, 2018). This threshold ensures a clear and meaningful distinction between long and short sentences in Austen and Doyle’s writing.

Pre-processing

- For sampling, the first 1000 sentences will be extracted from the text using a Python script. Sentences are defined as text strings ending with punctuation marks, such as ‘.’, ‘!’ or ‘?’. Sentences between apostrophes (””) will not be considered as a separate sentence.
- Then words will be counted. Words are defined as text strings in sentences delimited by spaces: ” ”.
- Then depending on the average number of words per sentence, the sentences will be characterized as short or long.

This data will be stored in a table as Table 1. Were the number of short and long sentences will be displayed per novel.

4 Predicted Results

Based on the previous literature, it is predicted that Austen’s novels will contain longer average sentence lengths than Doyle’s novels. This prediction is based on findings by Muchnik (2002), that longer sentences are indicative of elaborate narrative styles, which are more descriptive and focus on character development. Austen’s works are expected to reflect these traits, by containing elaborate sentences, with multiple clauses, that delve into social interactions and characters’ feelings. In contrast, Doyle’s novels are expected to have shorter sentences, consistent with action-driven narratives. These distributions can be seen in Table 2, where the expected distribution of sentence lengths, and their stylistics are illustrated. Matthews and Folivi (2023) argue that shorter sentences improve readability, this aligns with Doyle’s fast-paced and plot-driven writing style. His novels use clear, concise sentences, which keep the readers focused on the action and mystery.

Discussion The results are expected to illustrate how Austen and Doyle use sentence length to reflect their different narrative priorities. Austen’s

Author	Average sentence length	Stylistics
Jane Austin	Longer	Elaborate descriptions, character-driven
Arthur Conan Doyle	Shorter	Clarity, action-focused

Table 2: Distribution of sentence length and their stylistic implications

longer sentences would reflect her focus on complex characters and detailed narratives, attracting readers who enjoy in-depth, thoughtful stories. Doyle’s shorter sentences would support his goal of creating mystery and fast-moving plots for readers who value clarity and action. Shorter sentences help keep the attention, which is needed in detective fiction, where clarity and pacing are crucial in engaging the reader. This style fits the accessible approach that Matthews and Folivi (2023) describes, ensuring that Doyle’s works are both engaging and easy to follow. These findings can help us understand how different writing styles evolve to fit their genres and audiences. By comparing these styles, we learn more about how sentence length shapes storytelling.

5 Conclusion

By comparing the average sentence lengths in novels by Jane Austen and Arthur Conan Doyle, this study provides insight into how this sentence length reflects their distinct storytelling styles. Austen’s longer sentences focus on social interactions and character development, while Doyle’s shorter sentences match the more action-driven narrative, designed to keep the readers engaged. Future studies could apply these methods to other authors and expand this study to a larger corpus of fiction, it could examine how modern writers can use sentence length to target specific types of readers. Another possible research study could be examining the influence of historical and cultural factors on sentence structure.

GitHub Repository The code and data for this study are available at <https://github.com/daphnesmile-rug/Finalproject-Research-24.git>.

References

- Deveci, T. (2019). Sentence length in education research articles: A comparison between anglo-phone and turkish authors. *Linguistics Journal* 13(1), 73–100.
- Matthews, N. and F. Folivi (2023). Omit needless words: Sentence length perception. *PloS one* 18(2), e0282146.
- Muchnik, M. (2002). Sentence length in two novellas by judith katzir. *Hebrew Studies* 43, 7–20.
- Rudnicka, K. (2018). *Variation of sentence length across time and genre: Influence on syntactic usage in English*, pp. 219–240. John Benjamins Publishing Company.