

# Project Gutenberg Analysis

Shinjini Ghosh

July 11, 2020

## Analysis

1. *Title of the novel analysed:* The Mysterious Affair at Styles by Agatha Christie
2. *Total number of words:* 56410
3. *Total number of unique words:* 9925
4. *Top 20 most frequent words:*

| No. | Word | Count | No. | Word | Count |
|-----|------|-------|-----|------|-------|
| 1   | the  | 2624  | 11  | in   | 821   |
| 2   | i    | 1654  | 12  | he   | 718   |
| 3   | to   | 1336  | 13  | is   | 538   |
| 4   | of   | 1261  | 14  | not  | 511   |
| 5   | a    | 1201  | 15  | had  | 501   |
| 6   | and  | 1124  | 16  | she  | 482   |
| 7   | it   | 955   | 17  | her  | 462   |
| 8   | that | 938   | 18  | his  | 416   |
| 9   | was  | 926   | 19  | at   | 394   |
| 10  | you  | 848   | 20  | but  | 394   |

Table 1: Top 20 most frequent words

5. *Top 20 most frequent interesting words:* Here, we have excluded the top 1000 English words from consideration.

| No. | Word       | Count | No. | Word         | Count |
|-----|------------|-------|-----|--------------|-------|
| 1   | poiro      | 357   | 11  | howard       | 83    |
| 2   | inglethorp | 244   | 12  | asked        | 80    |
| 3   | mrs        | 231   | 13  | inglethorp's | 72    |
| 4   | mr         | 179   | 14  | mary         | 69    |
| 5   | john       | 161   | 15  | dorcas       | 69    |
| 6   | cavendish  | 137   | 16  | looked       | 66    |
| 7   | sir        | 106   | 17  | strychnine   | 65    |
| 8   | cynthia    | 98    | 18  | alfred       | 64    |
| 9   | into       | 95    | 19  | dr           | 61    |
| 10  | lawrence   | 84    | 20  | evidence     | 56    |

Table 2: Top 20 most frequent interesting words

6. *Top 20 least frequent words*, chosen randomly out of all those used just once:  
subsided, world-wide, attended, sensational, persist, invalidated, front, depressing, month's, sick, relations, senior, forty-five, yarn, inviting, again-after, seventy, autocratic, social, fondness
7. *Interesting word usage pattern*: Poirot comes into the book, not right at the beginning, but around the fifth chapter, and then has a *lot* of action before going a bit out of sight again towards the end. The number of usages of "Poirot" in each chapter of the book are [0, 1, 5, 1, 41, 79, 13, 49, 52, 17, 43, 18, 14].
8. *Look up a quote*: The quote "A man in love is a sorry spectacle." was found in Chapter 12.
9. *Randomly generated sentence*: We used a unigram model, randomly choosing a next word out of all the eligible words, to create a sentence starting with "The" and 20-word long. A few interesting sentences that came out are:
  - The jury is that?" "No." "Is it on the criminal or two, and wife." "Ah!" said John. "What does not
  - The old friend Japp was frowning perplexedly, and finally settled down the poisoned last moment—which partly your possession. No notice
  - The gardeners' evidence, but you must hurry to catch sight of the net he is correct, madame. But he goes,
  - The examination of the daily papers, all this is a moment's satisfaction of Hercule Poirot, and in which I dropped
  - The Scotland Yard fellows?" "Gone to substitute some important document—possibly a book?" "Yes, yes," I wonder?" "Some one to my