Universität Stuttgart
Institut für Literaturwissenschaft
Digital Humanities

A Digital Humanities Approach to Characterizing Murderers in Agatha Christie's Detective Novels

Masterarbeit zur Erlangung des akademischen Grades Master of Arts
der Philosophisch-Historischen Fakultät der Universität Stuttgart

vorgelegt von Mengjiao Li (B.A.)
Matrikelnummer: 3392731
E-Mail-Adresse: st183118@stud.uni-stuttgart.de

Erstgutachter: Dr. Anselm Küsters
Zweitgutachterin: Dr. Nora Ketschik

Einreichungsdatum: 04.12.2025

Abstract

This thesis employs methods in the field of Digital Humanities to analyze the characterization of murderers in eight detective novels by Agatha Christie. Moving beyond traditional close reading, the study integrates quantitative methods: social network analysis, speech/appearance rate visualization, Word2Vec contextualization, sentiment analysis, and linguistic representation, to uncover patterns in how male and female murderers are portrayed. The analysis reveals that Agatha Christie's murderers are diverse and lack distinct linguistic or contextual markers, quantitatively supporting the theoretical concept of the "unreadability" of her texts. Furthermore, the portrayal of female murderers demonstrates their narrative activeness, strategic use of societal stereotypes, and a blend of traditionally masculine and feminine traits, reflecting Agatha Christie's moderate feminism: the concept of the "New Woman." This research demonstrates how data-driven methodologies can complement and enrich traditional literary interpretation, offering new insights into Agatha Christie's narrative techniques and her enduring significance as an author.

Keywords: detective novels, Digital Humanities, Agatha Christie, murderers, unreadability, New Woman

Content

1. Introduction 1
2. Research Context 3
3. Operationalization 6
3.1 Concept and Definition 7
3.2 Corpus 8
4. Quantitative Analyses 11
4.1 Social Network Analysis (SNA) 12
4.1.1 Research Tradition 12
4.1.2 Annotation and SNA Procedure 13
4.1.3 Result of SNA 15
4.1.4 Synthesis of Results of SNA 37
4.1.5 Reflection of SNA 40
4.2 Bar Plot of Speech/Appearance Rate 42
4.3 Word2Vec Contextualization 44
4.3.1 Corpus Preparation and Model Training 45
4.3.2 t-SNE Visualization 47
4.3.3 Reflection of Word2Vec 49
4.4 Sentiment Analysis 50
4.4.1 Research Tradition 50
4.4.2 Identification of Modifiers 51
4.4.3 Result of Sentiment Analysis 54
4.4.4 Reflection of Sentiment Analysis 58
4.5 Linguistic Representation 59
4.5.1 Research Tradition 59
4.5.2 Result of Linguistic Representation 60
4.6 A Comprehensive Overview of the Murderers 66
5. Conclusion and Outlook 76
6. Literature 78
6.1 Primary Literature 78
6.2 Secondary Literature 78
6.3 Tools 80

6.4 Internet Access 81
6.5 List of Figures and Tables 81
6.6 Repository 83

1. Introduction

The "Golden Age of Detective Fiction" refers to the period of the 1920s and 1930s, during which the detective genre reached its peak. During this period, Agatha Christie emerged as one of its most prominent and influential authors. She defined the modern detective story, creating what is now known as the "whodunit" genre, in which the mysteries of murder are not unraveled until the very end. Agatha Christie wrote over 66 detective novels and short story collections, along with 14 plays, in addition to her other works. "When she died on 12 January 1976, she was known throughout the world as the Queen of Crime, unrivalled as the bestselling novelist of all time with two billion books sold in more than 100 languages." (*Agatha Christie an Autobiography,* cover text)

The two well-known and distinctive figures in Agatha Christie's novels are the two detectives: Hercule Poirot and Miss Marple. Hercule Poirot is a Belgian detective, often noted for his distinct Belgian accent and frequent use of French phrases, which adds to his uniqueness. Poirot is known for his self-assurance and pride in his intellect. He believes in the superiority of his "little grey cells" (his term for his powers of deduction). His confidence can be seen in his tendency to solve cases in a methodical and structured manner, often outsmarting both the police and the murderers. Miss Marple is a British woman, living in the fictional village of St. Mary Mead. Unlike Hercule Poirot, who is a professional detective, Miss Marple is an amateur sleuth, relying on her keen understanding of human nature. While Poirot is confident and takes pride in his intellect, Miss Marple remains humble, often underestimated by those around her.

Due to the distinct characterizations of these two detectives, a substantial body of research on them has emerged. There are analyses of these two detectives, focusing on their cognitive strategies, feminist interpretations of Miss Marple and so on. To name a few, Domjanović & Oklopčić (2022) analyze some of the murders portrayed in Agatha Christie's works, describe the methods used by Hercule Poirot and Miss Marple of solving the crimes, explain the main features of the two investigators, and compare them and their strategies of unravelling the mysteries surrounding the crimes; Najar & Vaziri (2021) explore the concept of criminal psychology and explicate its major tenets in the characterization of Hercule Poirot; Winterkvist (2020) compares Miss Marple's detective work with that of Sherlock Holmes, highlighting the gender discrimination and ageism Miss Marple faces,

and discussing how these factors influence her approach to solving crimes. The appreciation for Agatha Christie's deep psychological insight is validated by Zeiger (2013), who proposes Miss Marple's intuitive method as a valid model for psychological assessment, thereby bridging fictional narrative and clinical practice.

The primary method of character analysis in these studies is close reading, which emphasizes detailed, interpretive engagement with individual texts. While this approach offers nuanced insights, it often relies heavily on subjective judgment. In contrast, Digital Humanities offers a complementary perspective through distant reading, a data-driven method that identifies patterns across large collections of texts. The concept of distant reading is most associated with Franco Moretti, particularly in his article "Conjectures on World Literature." In this article, Moretti proposes an innovative reading method that goes beyond traditional literary canons, referring to the overlooked works as "the great unread." What distinguishes this approach in literary studies is its use of techniques such as sampling, statistics, paratexts, and other elements that are typically excluded from conventional literary analysis.

By incorporating distant reading alongside close reading, researchers can balance subjective interpretation with richer, quantitative evidence, thereby generating more comprehensive insights into literary characters. Scholars in the scholarship of Digital Humanities have increasingly advocated for such an integrated approach to enhance both qualitative and quantitative analysis. For instance, Jockers (2013) introduces the idea of macroanalysis, which involves analyzing large-scale literary data to uncover patterns that are not easily seen through traditional close reading and explores how computational techniques like distant reading can be combined with traditional close reading to deepen our understanding of literature. Similarly, Drakman et al. (2022) highlight how digital methods, especially through the combination of close and distant reading, enrich research in the history of science and ideas.

However, methods from Digital Humanities have been applied only sparingly to the study of Agatha Christie's novels. Justyna (2022) investigates links between particular texts to see whether they will create certain clusters representing e.g. genre, author or, in the case of Polish corpora, the translator's signal in her thesis "A stylometric analysis of detective fiction and other works of Agatha Christie, Arthur Conan Doyle and Gilbert Keith Chesterton." Theodar (2025) conducts a study to explore how cinematic adaptations of

Agatha Christie's Hercule Poirot novels, directed by Kenneth Branagh, reconfigure not only narrative structure but also the emotional and social systems embedded within them. The paper applies computational methods including sentiment analysis, character network modeling, and a custom-built Adaptation Divergence Index (ADI). The findings illuminate how Branagh's adaptations strategically humanize Poirot, elevate secondary characters, and shift genre logics from logical deduction to psychological horror.

The murderers in Agatha Christie's novels represent a significant lacuna in the scholarship. The characterizations of murderers, however, deserve closer examination, as they offer fresh and valuable insights into the characters and the author. Building on this opportunity, this thesis applies Digital Humanities methods to analyze the characterizations of murderers in eight of Agatha Christie's works. In doing so, it aims to contribute to a field that remains relatively underexplored.

This thesis is organized as follows: after the introduction, the second chapter outlines the research context with respect to Agatha Christie's murderers. The third chapter details the operationalization of the study, including definition of concepts and description of the corpus. The fourth chapter presents the quantitative analyses. A comprehensive overview of the murderers based on the results of these analyses, along with additional factors, is provided in the fifth chapter. Conclusion and suggestions for future research are presented finally.

2. Research Context

This chapter introduces studies that investigate the murderers in Agatha Christie's detective novels.

Ina Rae Hark (1997) argues that Agatha Christie's detective fiction both constructs and interrogates the illusion of readability: the idea that truth, guilt, and human behavior can be fully deciphered. Drawing on Peter Hühn's and Tzvetan Todorov's theories, Hark suggests that while detective fiction pretends to offer logical closure, its determinate meanings only mask the fundamental unreadability of human motives and morality. Hark contends that Agatha Christie's works dramatize this unreadability. Her impossible murderers: detectives, victims, or narrators who turn out to be killers, expose the limits of conventional reading

strategies. Stories like *The Murder of Roger Ackroyd*, *Murder on the Orient Express*, and *And Then There Were None* reveal that the reader's and detective's attempts to interpret signs of guilt are inevitably flawed, since guilt and innocence are intermixed rather than opposed. In *Curtain*, Hark finds Agatha Christie's most explicit statement of this philosophy. Poirot becomes both reader and author, decoding a textual murderer who manipulates others through language, and ultimately kills to stop further textual violence. Here Agatha Christie equates murder with acts of interpretation, implying that both are ways of imposing meaning on an unreadable world. Thus, for Hark, Agatha Christie's fiction turns detective reading into an allegory of textual and moral uncertainty. Beneath their surface puzzles, her novels acknowledge that true understanding of texts and of people, is impossible, making Agatha Christie a far more self-aware and subversive writer than her reputation for simplicity suggests.

Makinen (2006) analyzes the textual treatment of Agatha Christie's female villains and it reveals "a matter-of-fact acceptance of women as murderers," (Makinen 2006, 139) in a way that refuses to either demonize them for their rejection of gender stereotypes or to negate their agency and thus their power to disrupt society. Before she analyses the female murders in Agatha Christie's novels, she examined the representations and explanations for female violence current at the time of her fiction' s productions and "it took three forms: analysis of newspaper coverage of female murderers, popular books on female murderers and the current criminological explanation of female offenders." (Makinen 2006, 149) She concludes that "the fact is that Agatha Christie's representations of female murderers do not accord with the contemporary textual stereotypes of female misbehaviour, since they grant them a reasoned efficiency in their murderous plans in the novels which challenge dominant gender discourses, asserting the potential and powerful possibilities of transgressive femininity." (Makinen 2006, 149)

Verbogen (2008) analyzes patterns of murders in Agatha Christie's Miss Marple novels, focusing on the social class and gender of both murderers and victims. The study examines whether murderers belong to specific social classes, whether their motives vary by class and gender, and whether male and female murderers target victims differently. It also considers the victims' social classes and gender, highlighting that female victims outnumber male victims and exploring whether motives for murder differ by gender or class. Additionally, Verbogen investigates Agatha Christie's use of murder weapons, challenging the assumption that Agatha Christie primarily employs domestic tools, and

examines whether weapon choice correlates with the murderer's class or gender. The study concludes that while Agatha Christie employs formulaic elements, she balances them with inventive variations, ensuring originality and surprising readers. As a result, her novels cannot be dismissed as mere repetitions; her skill and creativity underpin her enduring reputation as the "Queen of Crime."

Dolski (2024) examines gender representation in Agatha Christie's detective fiction, specifically analyzing her male and female murderers to explore how they reflect and challenge traditional gender norms. It focuses on these murderers: male murderers Justice Wargrave (*And Then There Were None*) and Dr. James Sheppard (*The Murder of Roger Ackroyd*), and female murderers Jane Wilkinson (*Lord Edgware Dies*) and Bella Tanios (*Dumb Witness*). Dolski begins with an overview of crime and detective fiction, focusing on the Golden Age, and situates Agatha Christie within the historical and social context of Interwar Britain, noting shifts in gender roles after World War I. The analysis of male murderers applies Foucault's concept of power to examine how these characters exercise and represent power through crime and social behavior. It also uses Raewyn Connell's theory of masculinity and Judith Butler's gender theory to explore how male murderers conform to hegemonic masculinity and societal expectations of men. For female murderers, Dolski draws on Butler's idea of gender as performance and the concept of the "New Woman" to show how Agatha Christie's female criminals combine traditionally masculine traits with feminine qualities, challenging gender binaries and societal norms.

EverythingAgatha.com investigates the relationships between victims and murderers in Agatha Christie Poirot novels based on data analysis and statistics. "The data covers 82 victims and all murders mentioned in the books, all attempted murders and induced suicides. This study aims to address three research questions: (1) Is the world of Hercule Poirot more dangerous than that of Miss Marple? (2) Are men more frequently victims of murder than women in the Poirot series? (3) Does the method of murder or the gender of the victim show variation across the different narratives, in comparison to other works?" (*EverythingAgatha.com*, "Data Analysis Within the Poirot Novels", § 2) Several noteworthy findings emerge from the data analysis: female victims outnumber male victims, yet female murderers are more than male murderers; men are more likely to be murdered by either other men or women, whereas women are equally likely to be killed by either gender; in Poirot series, poison, likely influenced by Agatha Christie's wartime experience, emerges as the predominant method of murder, followed by stabbings.

Building on these frameworks and combining with digital methods, this thesis seeks to further investigate the gendered dimensions of murderers in Agatha Christie's works, focusing specifically on eight novels from the Poirot and Miss Marple series. By analyzing data on male and female murderers, with an emphasis on their relationships with other characters, their speech/appearance rate, their contextualization across the novels, sentiment analysis, and linguistic representation, this thesis explores whether the results align with or diverge from the findings of previous research of Agatha Christie's murderers and aims to assess whether Agatha Christie's conception of murderers and her moderate feminism are reflected in her works.

3. Operationalization

Pichler & Reiter (2022) present a model that outlines the workflow for a reflected text analysis (see Figure 1). This model serves as the foundation for the operationalization of this thesis.

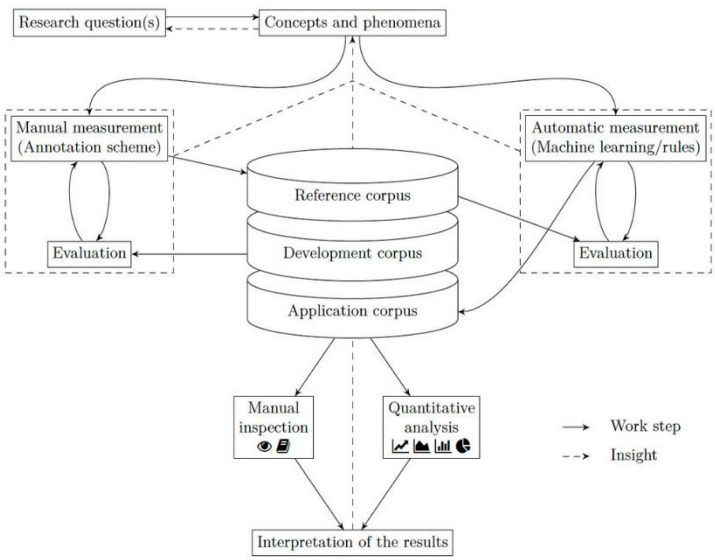

Figure 1: Workflow for Reflected Text Analysis, Pichler & Reiter (2022)

3.1 Concept and Definition

Figure 1 begins with the formulation of research questions that are closely linked to the underlying concepts and phenomena. The central research question of this thesis is how do the methodologies in Digital Humanities reflect the characterizations of the murderers through which Agatha Christie's conception of murderers and her moderate feminism can be revealed? To investigate this question, the theoretical concepts must be defined and translated into measurable units.

In this thesis, murderers' characterizations are reflected in these aspects: murderers' relationships with other characters, their talkativeness (reflected through speech/appearance rate), contextualization of murderers (semantic fields), sentiment of murderers' modifiers (whether they portrayed as positive, neutral or negative), and differentiation of modifiers for male and female murderers. These characterizations are investigated by employing social network analysis, bar plot visualization, Word2Vec analysis, sentiment analysis and linguistic representation.

Agatha Christie's conception of the murderers encompasses her approach to characterizing murderers. This concept is grounded in Ina Hark's theory of the unreadability of Agatha Christie's texts, which argues that "human guilt and innocence are thoroughly intermixed, rather than being in mutually exclusive binary positions. If we attempt to read people like books to separate the murderous from the non-murderous, we must fail, because in Christie's view, no such creature as a person incapable of murder exists." (Ina Hark 1997, 2) It also draws upon Verbogen (2008)'s argument that, while Christie employs formulaic structures, her works transcend simple repetition through inventive variations. Combining the results from the above-mentioned methodologies in Digital Humanities with factors such as the murderers' motives, killing methods, fate and so on into their portrayals, this thesis reveals Agatha Christie's conception of the murderers: that anyone has the potential to commit murder. Furthermore, it highlights her inventive variations in constructing these characters.

Feminism in Agatha Christie's detective novels is prominently represented through her portrayals of female murderers. Dolski (2024) applies the concept of the "New Woman" to analyze the female murderers in Agatha Christie's works. The "New Woman" is "a woman who represented women seeking greater independence, education and involvement in

public life, challenging traditional gender roles and societal expectations." (Dolski 2024, 33) In Agatha Christie's novels, the female murderers are not submissive, nor are they oppressed by men or forced into crime. Rather, they are women with their own criminal intent, who are aware of the roles that society assigns to them and the expectations others have of them. They are determined and intelligent women who understand how to exploit social expectations to manipulate situations, mislead investigations, and deflect suspicion. By examining the differences between male and female murderers based on the results of quantitative analyses, this thesis aims to uncover how the concept of the "New Woman" is reflected.

3.2 Corpus

The corpus of this study consists of eight detective novels by Agatha Christie. To distinguish between female and male murderers, three novels featuring male murderers are selected: *The Murder of Roger Ackroyd* (1926) with Dr. James Sheppard as the murderer, *The A.B.C. Murders* (1936) with Franklin Clarke as the murderer, and *And Then There Were None* (1939) with Justice Wargrave as the murderer. Three novels featuring female murderers are selected.: *Lord Edgware Dies* (1933) with Jane Wilkinson as the murderer, *A Murder is Announced* (1950) with Miss Blacklog as the murderer, and *The Mirror Cracked from Side to Side* (1962) with Marina Gregg as the murderer. The final two novels each features both male and female murderers, offering an opportunity to investigate similarities and distinctions of female and male murderers when they are placed in the same story: *The Mystery of the Blue Train* (1928) with Major Knighton and Ada Mason as the co-murderers, and *Evil Under the Sun* (1941) with Patrick Redfern and Christine Redfern as the co-murderers.

Given the limited scope and timeframe of this thesis, only eight out of 66 detective novels of Agatha Christie are selected. The selection of these eight novels is motivated by methodology used in statistical sampling, where random selection helps mitigate bias and ensures that the novels analyzed offer a cross-section of Agatha Christie's writing across different periods: from the 1920s to the 1960s. This chronological span also allows for an examination of whether Agatha Christie's view on feminism has changed overtime.

The total number of tokens across the eight novels is 805,967. Novels featuring male solo murderers account for 256,197 tokens, representing approximately 31.9% of the total, while those featuring female solo murderers comprise 301,554 tokens, representing about 37.6% of the total. With respect to the novels featuring co-murderers: *Evil Under the Sun* contains 148,816 tokens (18.5% of the total) and *The Mystery of the Blue Train* contains 99,400 tokens (12.4% of the total). Overall, the distribution of tokens across novels and murderer gender is relatively balanced, providing a robust basis for comparative analysis. These novels can be freely accessed on Project Gutenberg (https://www.gutenberg.org) and in Classical Detective: The Poetics of the Genre (https://detective.gumer.info/english.html). Below are plot overviews of these eight selected novels. [1]

The Murder of Roger Ackroyd The murderer in this novel, Dr. James Sheppard, is also the narrator. He lives with his sister Caroline in the village of King's Abbot. The wealthy widow Mrs. Ferrars commits suicide after confessing to Roger Ackroyd that she poisoned her husband and is being blackmailed. That same night, Ackroyd is found murdered in his study, and suspicion falls on his stepson Ralph Paton, who mysteriously disappears. Hercule Poirot is asked to investigate. Ultimately, Poirot reveals that the real murderer is Dr. Sheppard himself, who also blackmailed Mrs. Ferrars. Sheppard kills Ackroyd to prevent exposure, covers his tracks with a dictaphone recording, and omits incriminating details from his narration. Poirot advises Sheppard to take his own life to spare Caroline shame. The novel ends with Sheppard's written confession, which is the very story the reader has just finished.

The A.B.C. Murders Poirot receives letters signed "A.B.C." warning of murders committed in alphabetical order, with each crime scene marked by an ABC Railway Guide. After three such killings, suspicion falls on Alexander Bonaparte Cust, an epileptic salesman who fears he might be guilty but has solid alibis. Poirot proves that Cust is being framed and exposes the true culprit: Franklin Clarke, who murders his brother Sir Carmichael to secure his inheritance and disguises it as part of a serial killing spree. When confronted, Franklin tries to kill himself but fails, and Cust is cleared of all blame.

And Then There were None Ten strangers are invited to a secluded island by a mysterious "Mr. and Mrs. Owen." A gramophone recording accuses each of them of being responsible for someone's death that went unpunished. Soon, the guests are killed one by one in ways

[1] The plot overviews are summarized by the author based on the plot summary in Wikipedia.

that mirror a nursery rhyme, with a figurine removed from the dining table after each death. When the police arrive, they find no survivors and no clear culprit. Only a posthumous confession in a bottle reveals the truth: Judge Wargrave masterminds the scheme, driven by both a passion for justice and a hidden bloodlust. He arranges the invitations, manipulates Dr. Armstrong to help fake his own death, murders the others in sequence, and finally kills himself to make the mystery appear unsolvable.

Lord Edgware Dies Jane Wilkinson asks Poirot to help her divorce Lord Edgware, but he discovers that Edgware has already agreed. The next day, Edgware is found murdered. Although witnesses place Wilkinson at the scene, she also has an alibi from a dinner party. Poirot suspects that Carlotta Adams, a skilled impressionist, was hired to impersonate Wilkinson. However, Adams is soon found dead from an overdose. Evidence and a tampered letter shift suspicion onto others, but Poirot ultimately proves that Wilkinson kills her husband so she can marry the Duke of Merton, who will not wed a divorcée. She then kills Adams to silence her and later murders Donald Ross when he notices inconsistencies. Poirot exposes her lies, and Wilkinson is arrested, showing no anger or remorse for her crimes.

A Murder is Announced A newspaper advertises that a murder will occur at Little Paddocks, the home of Miss Letitia Blacklog. That evening, a man bursts in with a torch, shots are fired, and he is later found dead, identified as Rudi Scherz, a petty criminal. Inspector Craddock and Miss Marple investigate, suspecting that the real target is Miss Blacklog, who stands to inherit a fortune. Soon after, her companion Dora is poisoned, and another Miss Murgatroyd is strangled. Miss Marple uncovers the truth: the real Letitia died years earlier, and her sister Charlotte Blacklog assumed her identity to claim the inheritance. Scherz had recognized her from Switzerland, so she staged the holdup and killed him, then silenced Dora and Murgatroyd when they grew suspicious. Charlotte Blacklog's disguise unravels when Miss Marple sets a trap, leading to her breakdown and arrest.

The Mirror Cracked Side from Side Film star Marina Gregg and her husband Jason Rudd host a fête at their new home, Gossington Hall. During the party, Heather Badcock drinks from Marina's cocktail after accidentally spilling her own and suddenly dies. Two more deaths follow: Marina's secretary is poisoned, and her butler is shot. At first, it seems Marina is the target, but Miss Marple uncovers the truth: Marina kills Heather. Years earlier, Heather, unknowingly ill with German measles, had insisted on meeting Marina while she

was pregnant. Marina contracted the illness, resulting in her child's disability and her own breakdown. When she encounters Heather again, Marina poisons her in a sudden act of revenge. To cover her crime, she stages threats against herself and kills others who grow suspicious. In the end, Marina dies from an overdose, apparently administered by Jason to prevent her from killing again.

The Mystery of the Blue Train Onboard the Blue Train to the French Riviera, Ruth Kettering is found strangled in her compartment, and her valuable ruby, the "Heart of Fire," is missing. Suspicion falls on her husband, Derek Kettering, but Poirot investigates carefully. He discovers that Ruth's maid, Ada Mason, is actually the actress Kitty Kidd, and that the murder and jewel theft are orchestrated by Major Knighton, Ruth's father's secretary, who is secretly the notorious jewel thief Le Marquis. Mason helps Knighton gain access and create false leads. Derek has only briefly entered Ruth's compartment and is innocent. Poirot exposes Knighton and Mason's involvement, solving both the murder and the theft.

Evil under the Sun While on holiday at a Devon hotel, Hercule Poirot notices the flirtatious Arlena Marshall, married to Kenneth Marshall, drawing attention from Patrick Redfern. Arlena is later found strangled at Pixy Cove. Investigation reveals a carefully orchestrated plot: Patrick and his wife Christine Redfern conspire to kill Arlena to conceal Patrick's embezzlement of her fortune. Christine manipulates timelines and uses disguises to mislead witnesses, while Patrick commits the murder in the cave. Poirot uncovers their scheme, including the false alibis and staged evidence. Christine also provokes Arlena's stepdaughter, Linda, into a suicide attempt to further cover their tracks. Both Patrick and Christine are exposed, and Poirot reassures Linda that she is never responsible for the murder.

4. Quantitative Analyses

This chapter focuses on the quantitative analyses in this thesis, representing the methods and tools employed, results of quantitative analyses, and reflection of the analyses.

4.1 Social Network Analysis (SNA)

The first quantitative method used in this thesis is social network analysis (SNA).

4.1.1 Research Tradition

"Network analysis in literary studies was driven by a use case for social network analysis of a computer scientist Donal Knuth, who needed example data for the Stanford Graph Base, a program and dataset collection for the generation and manipulation of graphs and networks. The list of datasets featured character interactions in the chapters of *Anna Karenina*, *David Copperfield*, and *Les Misérables*. The files for these three novels contain data on the co-occurrence of literary characters per chapter, which makes for genuine network data. After some more individual approaches to the network analysis of novels, it took some more years until network analysis in literary studies eventually happened with the studies of Franco Moretti in 2011 and Peer Trilcke in 2013." (Fischer et al. 2021, 519-520) Moretti's "Distant Reading" (2011) explores how network analysis can be used to study literary history and the relationships between authors, texts, and genres. In the article "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft," Peer Trilcke provides a more empirical and data-driven approach to study literature and to reveal patterns and structures that traditional literary analysis might overlook.

SNA's main applications in literary studies are: (1) analysis of relationships between characters in literature: through a visualization of the relationships and metric analysis, the most central and important character can be identified, dynamics understood und hidden structure of the literature uncovered. For example, Agarwal et al. (2012) investigate the relationships and interactions between characters in Lewis Carroll's *Alice's Adventures in Wonderland* using network analysis, mapping out and examining the network of characters to uncover patterns, central figures, and the overall structure of their interactions within the story. Stiller et al. (2003) demonstrate through SNA that network structures in Shakespear's drama exhibit small-world properties and resemble socially studied societal structures; (2) visualization of correspondences, contacts, change of living places and other interactions of authors: this reveals their social and geographical positions and the roles they play within a network of literary contacts. Görmar et al. (2003) examine the social network that the scholar Sturm built through contacts, loans, correspondences, and other interactions. This study combines digital and hermeneutic methods to comprehensively

analyze the intellectual networks of the early modern Republic of Letters; (3) the mutual/unilateral influence of literary works: such network analysis provides insights into the relationships between texts: the intertextuality. For example, Winckles et al. (2017) argue that networks not only provided women with access to the literary marketplace, but fundamentally altered how they related to each other, to their literary production, and to the broader social sphere; (4) comparing writing styles of different authors: Eder (2015) discusses reliability issues of a few techniques used in stylometry and demonstrates that the best way of extending cluster analysis dendrograms with a self-validating procedure is using a new visualization technique, which combines the idea of the nearest neighborhood derived from cluster analysis, the idea of hammering out a clustering consensus from bootstrap consensus trees, with the idea of mapping textual similarities onto a form of a network.

In this thesis, SNA is applied to represent the relationships between each murderer and the other characters in the eight selected novels.

4.1.2 Annotation and SNA Procedure

A network consists of nodes and the connections that link them. Nodes, also referred to as vertices, are entities, like figures, places. (Schumacher 2018) Connecting lines, also referred to as edges, are the relationships between nodes. Edges in a network can be directed or undirected. (Jansen 2003, capital 5-6) Besides nodes and edges, there are another two key components in SNA: metrics and visualization. Metrics are the quantitative measures to analyze a network, and visualization is a representation of the network in graphic.

In this thesis, the relationships between murderers and the others are based on co-occurrence. When characters appear together in a chapter, they have a relationship of co-occurrence. So, the nodes are characters and the edges represent co-occurrences. Co-occurrences in this study are manually recorded since characters who are merely mentioned, never actively participating in the narrative or who are already deceased are excluded from co-occurrence calculations. Consequently, the use of Named Entity Recognition (NER) could have introduced irrelevant results, and is therefore not adopted.

After co-occurrences are recorded, the data is inputted into Ezlinavis (Fischer et al. 2017), a semi-automatic software, to generate a CSV file containing *source*, *type*, and *target*, for a subsequent analysis and visualization in a network analysis tool. There are plenty of tools for visualization and the most common used tool in Digital Humanities is Gephi (Bastian et al. 2009). It is well-suited for addressing questions about network-like constellations and structures. Gephi also provides a user-friendly interface and different algorithms, namely different layouts, which provide different methods for positioning nodes and visualizing relationships in a network. Thus, after the CSV file is generated, it is imported into Gephi and a co-occurrence-based network is initiated.

To elucidate distinct characterizations of the murderers, several features of characters are added as attributes in Gephi's *Data Labor*. These features are received through annotation. To be annotated are the quoted speeches of characters as well as their mentions by others. Quoted Speeches[2] refers to the spoken sentences of each active character, which are labeled with their names. Since quoted speeches indicate interactivities of characters, the number of it is a sign whether the character is active throughout the novels. Not to be annotated are then quoted sentences which are thoughts and monologues, as these do not reflect interactive dialogues. When a character speaks to another person and mentions a third character in their dialogues, the third character should be annotated as Mention[3], labeled as *mention_name*. However, if the third character is aware that he/she is being mentioned, the third character should not be annotated. For example, if character A introduces character C to character B, C should not be annotated, as C knows he/she is being mentioned. A mention indicates significant attention or suspicion directed toward a character. Besides, mention of victims is not annotated. Even if a victim is initially under strong suspicion, their death removes that suspicion, so their mentions are no longer considered. An exception applies to *And Then There Were None*: in this novel, all ten individuals are found dead including the murderer Justice Wargrave, and since there is neither a detective nor an inspector, all victims are at some point suspected as the possible murderer. Therefore, mentions of these victims should still be annotated.

In this thesis, manual annotation is employed, as there are several reasons: (1) quoted monologues and internal thoughts, often marked with quotation marks, are not to be

[2] Quoted Speeches is capitalized when it refers to a metric in SNA.

[3] Mention is capitalized when it refers to a metric in SNA.

annotated and therefore require manual verification after any automated process; (2) across the eight novels, quotation conventions vary: in some, speeches are enclosed in single quotation marks, while others use double quotation marks, such as *“”* or ″ ″; (3) certain phrases or words appear in quotation marks but are not actual quoted speeches, necessitating manual discernment; (4) it remains challenging for current computational methods to accurately attribute quoted speeches to the correct speaker..

These eight selected novels by Agatha Christie are manually annotated with CATMA 7 (Gius et al. 2024). CATMA (Computer Assisted Text Markup and Analysis) is a web-based tool designed for the manual and semi-automatic annotation, analysis, and visualization of texts. It allows for the creation of custom tag sets as well as collaborative annotation. (Schumacher, 2019b) In this study, the two tags are names of the characters (for Quoted Speeches) and *mention_(character name)* (for Mention). Since the to be annotated quoted speeches and mentions are obvious, there is no need for more than one annotator to calculate the IAA (Inter-Annotator Agreement) and to generate a gold standard.

4.1.3 Result of SNA

The result of SNA is presented below, with three distinct networks corresponding to three different rankings for each novel: Degree, Quoted Speeches, and Mention. In addition to these three measurements, three centrality metrics: Closeness Centrality, Betweenness Centrality, and Eigenvector Centrality, are also examined.[4] Prior to discussing the results, it is important to clarify the definitions of the centrality measurements, as the metrics related to Quoted Speeches and Mention have already been defined in 4.1.2.

Degree, is also called Degree Centrality. “The simplest definition of actor centrality is that central actors must be the most active in the sense that they have the most ties to other actors in the network or graph.” (Wasserman & Faust 1994,178) A node with a high Degree Centrality is considered important because it has many direct connections to other nodes. Closeness Centrality is based on closeness or distance. “The measure focuses on how close an actor is to all the other actors in the set of actors. The idea is that an actor is central if it can quickly interact with all others.” (Wasserman & Faust 1994, 183) In a network, a node

[4] Degree/Degree Centrality, Closeness Centrality, Betweenness Centrality and Eigenvector Centrality are capitalized when they refer to the metrics in SNA.

with a high Closeness Centrality is closely related to other nodes and can spread information very fast. Betweenness Centrality refers to those actors who lie on the shortest path between other actors. A node with high Betweenness Centrality acts as a bridge in the network. It controls information flow between other nodes. (Wasserman & Faust 1994) Eigenvector Centrality relates to the concept of social capital defined by Bourdieu (1983). Bourdieu says that the volume of social capital is a function of two things: (1) the number of social connections possessed by a given agent; (2) the amount of social (and other) capital held by those with whom connections are maintained. Eigenvector Centrality means if a node has many relations and those relation partners have themselves many relations, then that node will have high eigenvector centrality, i.e., social capital.

The Murder of Roger Ackroyd

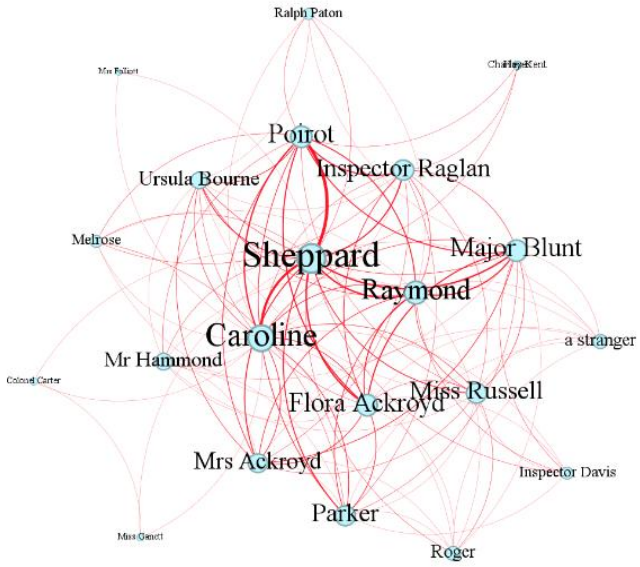

Figure 2: Visualization of Character Relationships in *The Murder of Roger Ackroyd* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 2 is a visualization of character relationships with nodes ranked by Degree. The node of Sheppard, who is the murderer and the narrator, is the largest based on co-occurrence, since he appears in every chapter. The other large nodes, Caroline (Sheppard's sister), Hercule Poirot, Flora Ackroyd (Roger Ackroyd's daughter), Parker (the butler),

Major Blunt (a friend of Roger Ackroyd and is in love with Flora Ackroyd), and Raymond (Roger Ackroyd's secretary) are also closely related to each other, as they are consistently present during the interrogation and investigation process. Since Sheppard is the narrator, on co-occurrence-based Degree Centrality is insufficient to fully investigate his characterization as a murderer. Additional metrics are therefore required. Figure 3 is a visualization with nodes ranked by Quoted Speeches.

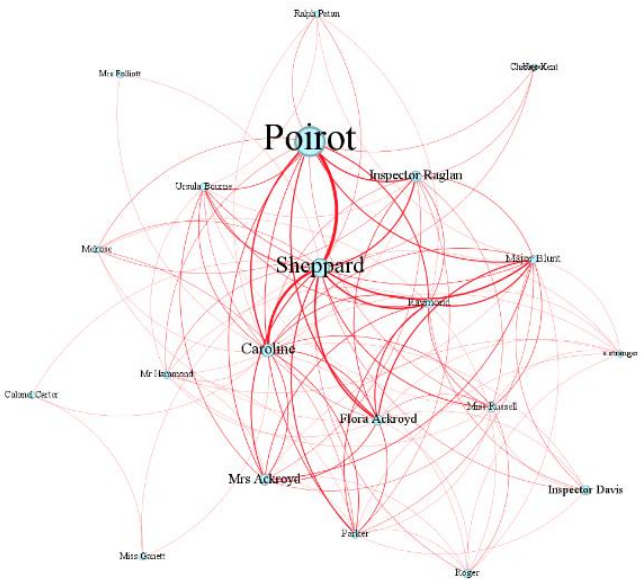

Figure 3: Visualization of Character Relationships in *The Murder of Roger Ackroyd* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

The node of Sheppard in Figure 3 is only smaller than that of Poirot, indicating that as a narrator he also interacts and talks a lot with other characters throughout the novel. Figure 4 is a visualization with nodes ranked by Mention, highlighting a feature associated with suspects or attention in a murder investigation. In Figure 4, the node representing Sheppard is barely visible, and his Mention value is indeed 0. This results from the annotation rule that a character is not annotated as Mention when he or she is aware of being referred to by others. Since Sheppard appears in every chapter, each mention of him occurs in his presence, and therefore none of these instances qualify as Mention under the annotation criteria. In contrast, the node of Ralph Paton becomes the largest, highlighting his prominence as the most suspicious character in the novel. Since Ralph Paton is indeed the

most suspicious person in the story (due to his disappearance on the crime night, his need for money and his bad relationship with Roger Ackroyd), this result of quantitative analysis aligns with the result from close reading.

Sheppard's Closeness Centrality (1.0), Betweenness Centrality (48.69), and Eigenvector Centrality (1.0) are the highest. It is understandable, as he is the narrator and appears in every chapter, he is the one who can spread information most rapidly and connect all other characters. As a private doctor of Roger Ackroyd and an "assistant" to Poirot, both of whom possess considerable social capital, Sheppard's Eigenvector Centrality is the highest.

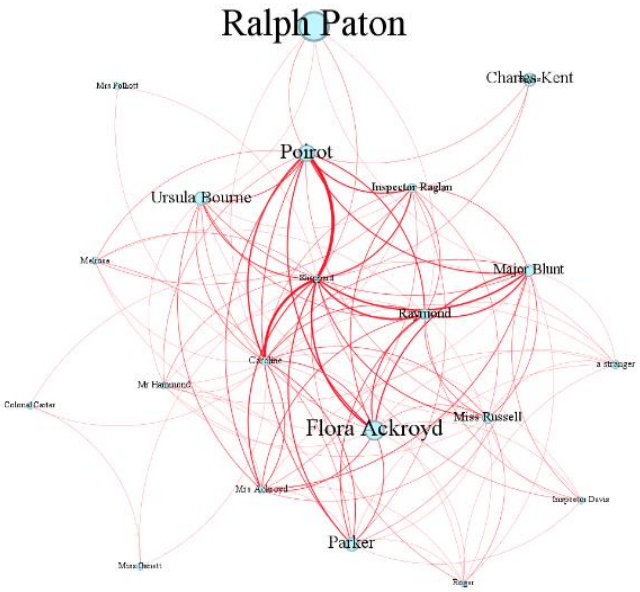

Figure 4: Visualization of Character Relationships in *The Murder of Roger Ackroyd* with Nodes Ranked by Mention, Layout: Yifan Hu

The A.B.C. Murders

In *The A.B.C. Murders*, Hastings serves as the narrator, and wherever he appears, Poirot is also present. As a result, the nodes representing Hastings and Poirot, ranked by Degree, are undoubtedly the largest (see Figure 5). Franklin Clarke, although introduced relatively late in the novel, has a Degree Centrality value of 11 (the highest value is 37). This is nearly as prominent as Cust's, who is a key suspect with a value of 14. In the visualization with

nodes ranked by Quoted Speeches (see Figure 6), Franklin Clarke's Quoted Speeches, with a value of 229, is only slightly lower than that of the narrator Hastings (289) and Inspector Crome (295), despite Franklin Clarke's later presence in the novel. The character with the highest value of Quoted Speeches is Poirot, with a value of 1557, followed by Inspector Crome in the second place. This result suggests that Franklin Clarke is an active character in the novel, as he speaks a lot.

According to Figure 7, Franklin Clarke's Mention is not prominent, since he has shifted suspect onto Cust, who is the most mentioned character and with the highest value of 178. The prominent nodes following Cust are Franz Ascher (60), Donald Fraser (40), and Miss Grey (57), who are also suspects in the murders of the first, second, and third victims, respectively. Franz Ascher is the estranged, alcoholic husband of the first victim, Mrs. Ascher, and he is considered to be the most likely person to have committed the complicated crime, especially given his lack of a solid alibi; Donal Fraser has a strained relationship with Betty, the second victim, and has even threatened to kill her; Miss Grey's strong desire for the Clarke family's fortune put her in a position to benefit from the murder of the third victim, Sir Carmichael Clarke. This result validates again that Mention is an efficient indicator of suspect or attention.

Franklin Clarke's Closeness Centrality has a value of 0.54, following the highest value of Poirot (0.81) and Hastings (0.81). This suggests that he is relatively well-connected to other characters, including members of his own family and those involved in the investigation. Since Franklin Clarke appears late in the novel and his appearance is limited in the family Clarke, he does not play a role in connecting different social groups, thus, his Betweenness Centrality is low. Franklin Clarke's Eigenvector Centrality is at an above-average level, with a value of 0.47, following the second highest value of 0.48 (the highest value is 1), reflecting his association with influential characters such as Hercule Poirot and Hastings.

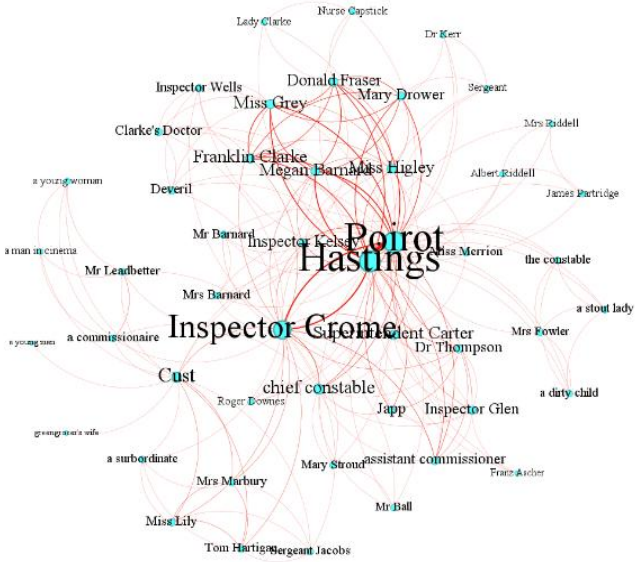

Figure 5: Visualization of Character Relationships in *The A.B.C. Murders* with Nodes Ranked by Degree, Layout: Yifan Hu

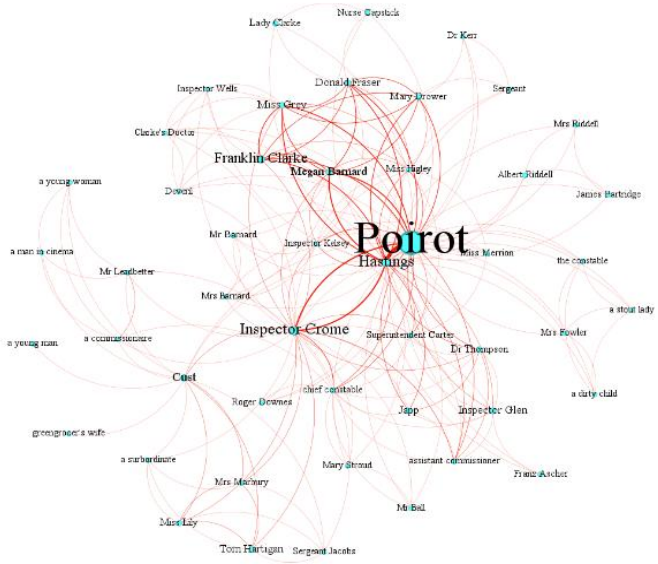

Figure 6: Visualization of Character Relationships in *The A.B.C. Murders* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

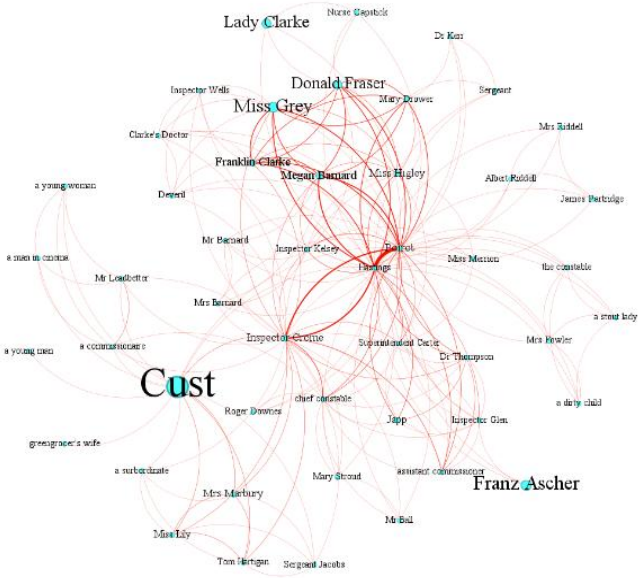

Figure 7: Visualization of Character Relationships in *The A.B.C. Murders* with Nodes Ranked by Mention, Layout: Yifan Hu

And Then There were None

In *And Then There Were None*, there is neither detective nor inspector. Vera Claythorne and Philip Lombard are the two characters who die last in the novel, giving them the largest nodes ranked by Degree (see Figure 8). Although Justice Wargrave commits suicide after the death of Vera and Lombard, he has already staged his own death prior to them. His culpability is revealed only through his posthumous confession in a letter, meaning that, within the plot's temporal sequence, his active presence ends before that of Vera and Lombard. The sizes of nodes ranked by quoted speeches also correspond to the death sequence; therefore, while Justice Wargrave's node is not the largest, it still holds a prominent position. He leverages his identity as a former judge to gain the others' trust and to mislead them through his authoritative and professional manner of speaking (see Figure 9).

Justice Wargrave is not frequently mentioned throughout the novel, as his status as a retired judge earns him the trust of the others, leaving no one to suspect him (see Figure 10). When a murder occurs, he assumes the role of leading the group in analyzing the case and

proposing subsequent courses of action. Dr. Armstrong and Rogers have the largest nodes when ranked by Mention. Dr. Armstrong is frequently referred to because he is considered the most likely person to have access to poison which is the cause of Mrs. Rogers's death, while Rogers is suspected of knowing the identity of "Mr. and Mrs. Owen," owing to his position as the butler.

Lombard has the highest Closeness Centrality (1.0), Betweenness Centrality (16.3), and Eigenvector Centrality (1.0), followed by Justice Wargrave, Dr. Armstrong, Vera, and Emily Brent, all of whom share identical centrality values of 0.93, 3.3, and 0.99, respectively. Although Justice Wargrave "died" before Dr. Armstrong, Vera, Emily Brent, and Lombard, his Closeness, Betweenness, and Eigenvector Centralities are comparable to theirs. This suggests that while he is alive, he maintains close relationships with the other characters, is related to the influential characters, and occupies a pivotal, central role in the narrative, reflecting his influence and authority as a former judge.

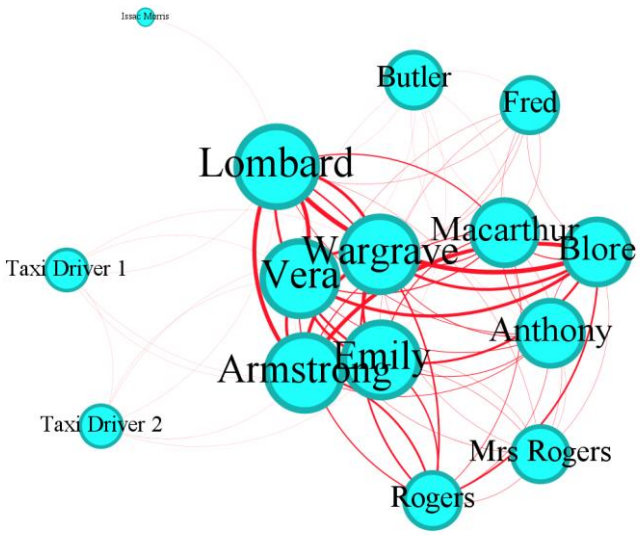

Figure 8: Visualization of Character Relationships in *And Then There were None* with Nodes Ranked by Degree, Layout: Yifan Hu

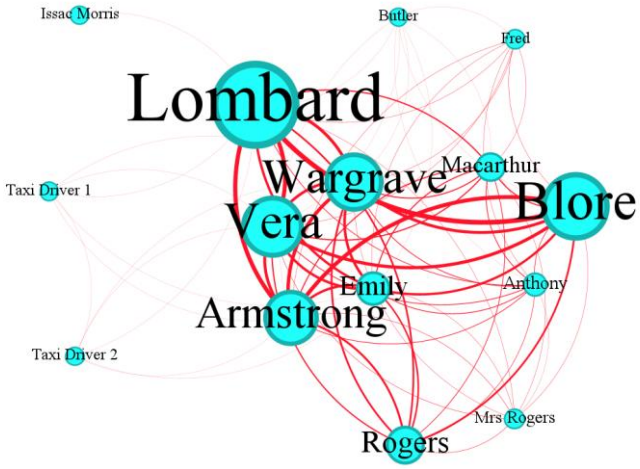

Figure 9: Visualization of Character Relationships in *And Then There were None* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

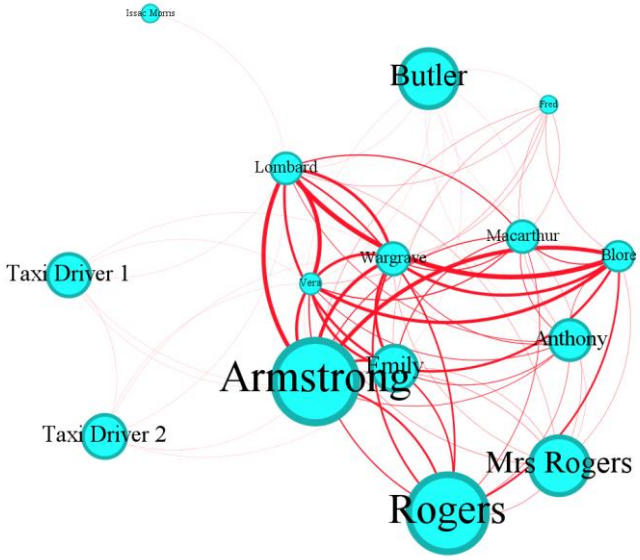

Figure 10: Visualization of Character Relationships in *And Then There were None* with Nodes Ranked by Mention, Layout: Yifan Hu

Lord Edgware Dies

Hastings is again the narrator of *Lord Edgware Dies* and whenever and wherever he appears throughout the novel, Poirot is with him. Thus, the biggest nodes ranked by Degree are undoubtedly Hastings and Poirot (see Figure 11). As the murderer, Jane Wilkinson does not appear as frequently as Hastings or Poirot; her appearances are concentrated mainly in the early chapters, yet they are still relatively high compared with most other characters.

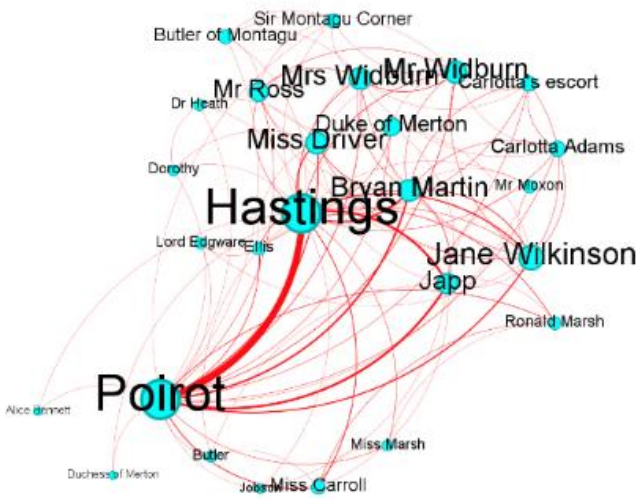

Figure 11: Visualization of Character Relationships in *Lord Edgware Dies* with Nodes Ranked by Degree, Layout: Yifan Hu

In Figure 12, Poirot's node is particularly prominent. As a detective, he engages in extensive communication. The nodes of the other characters are considerably smaller compared with Poirot's. Among them, Jane Wilkinson's node stands out at a noticeable level, suggesting that she is relatively talkative in the narrative and it is indeed certain statements of her that redirect suspicion onto herself.

Jane Wilkinson's node of Mention is the largest (see Figure 13). Jane Wilkinson is mentioned more frequently in the novel because her actions and conversations create multiple points of suspicion. Her request to Poirot regarding a divorce with Lord Edgware,

the impersonation of her by Carlotta Adams (the second victim) to establish an alibi, and the contradictions in her statements all draw attention to her, forcing Poirot to mention her repeatedly during the investigation. Consequently, she occupies a prominent position in the narrative despite her relatively limited direct involvement in the central plot.

Jane Wilkinson's Closeness Centrality (0.69) is lower than that of Poirot and Hastings, each with a value of 1.0, suggesting her relatively close relation to other characters. Her connections stem from multiple facets: her role as Mrs. Edgeware, her personal request to Poirot, her private connection with Carlotta Adams, her romantic involvement with Duke of Merton, her friendship with Bryan Martin, her social interactions with Mr. Ross (the third victim), among others. Betweenness Centrality of Jane Wilkinson (9.4) is much lower than that of Poirot and Hastings (88.8), suggesting that while she is connected to many characters, she rarely serves as an intermediary in interactions between others.

Due to her connections with influential characters in the novel, such as Poirot, Bryan Marin, Jane Wilkinson's Eigenvector Centrality (0.75) is slightly lower than that of Poirot and Hastings (1.0) but at a high level.

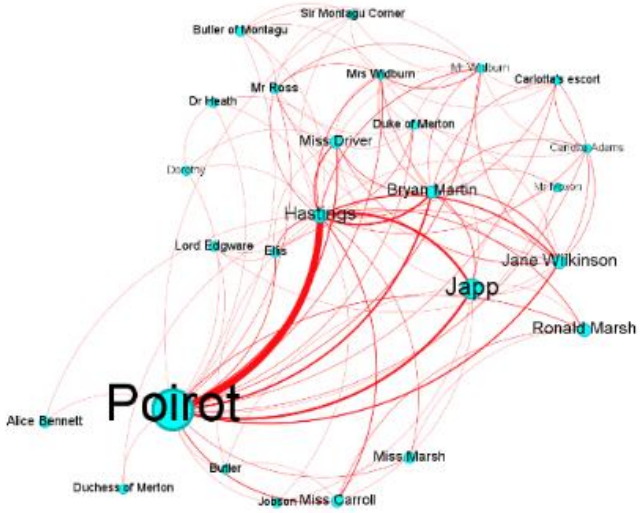

Figure 12: Visualization of Character Relationships in *Lord Edgware Dies* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

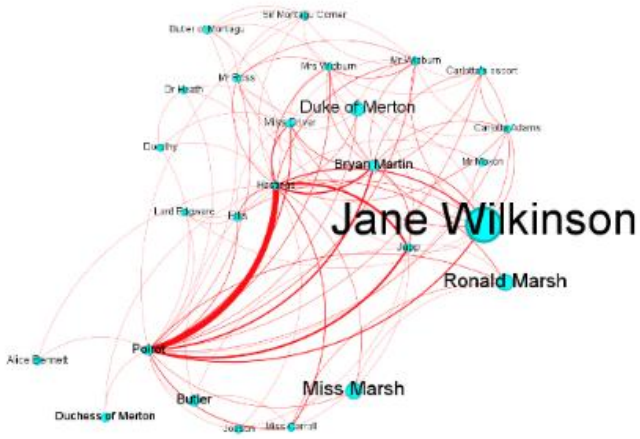

Figure 13: Visualization of Character Relationships in *Lord Edgware Dies* with Nodes Ranked by Mention, Layout: Yifan Hu

A Murder is Announced

In *A Murder is Announced*, the node representing the murderer, Miss Blacklog, is smaller than the largest node, that of Inspector Craddock (see Figure 14). It is, however, nearly the same size as the other nodes of those who were present at the party where the murder occurred, indicating that Miss Blacklog shares a strong connection with the other individuals involved.

With respect to Quoted Speeches (see Figure 15), Miss Blacklog's node, with a value of 647, is slightly smaller that of Inspector Craddock (742) and larger than that of Miss Marple (368), indicating that Miss Blacklog is an interactive character with a notable level of engagement in the narrative.

In the visualization of character relationships with nodes ranked by Mention (see Figure 16), Miss Blacklog's node is the largest one. Miss Blacklog creates the illusion of a death party, leading others to believe that the true target of the murderer was her. However, she could not figure out who would have the motive to kill her, which results in the absence of any suspects who could be considered as the real killer. Thus, the number of Miss

Blacklog's Mention is an important indicator of her role as a suspect: a clue that becomes visible only through quantitative analysis It is precisely because she is mentioned so frequently, and because she fears her secrets may be revealed, that she kills Dora, her closest friend. Furthermore, as suspicion around her begins to grow, she proceeds to eliminate the third victim.

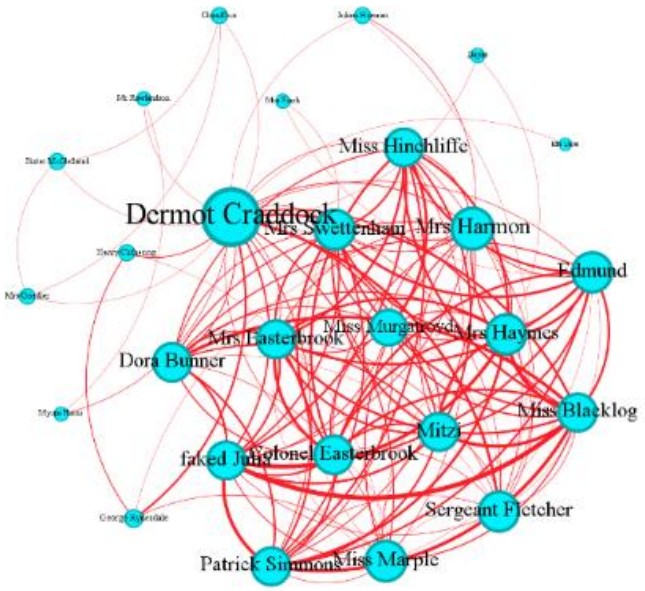

Figure 14: Visualization of Character Relationships in *A Murder is Announced* with Nodes Ranked by Degree, Layout: Yifan Hu

Miss Blacklog ranks third in terms of Closeness Centrality, with a value of 0.7, following Inspector Craddock (0.96) and Miss Marple (0.72). Interestingly, several other characters who were present at the party also exhibit Closeness Centrality values similar to that of Miss Blacklog. This suggests that, while Miss Blacklog is well-connected within the network, her position is not significantly more central than that of many others. Miss Blacklog's Betweenness Centrality (11.16) and Eigenvector Centrality (0.89) are also neither particularly high nor low in comparison to other characters in the network.

Overall, Miss Blacklog maintains a neutral position within the narrative structure, presenting herself as an ordinary, largely unsuspected elderly woman. Making other characters believe the attempted murder at the party was directed at her, she effectively

diverts suspicion away from her true role in the crime. Her suspect is only uncovered through quantitative analysis, particularly through the metric of Mention.

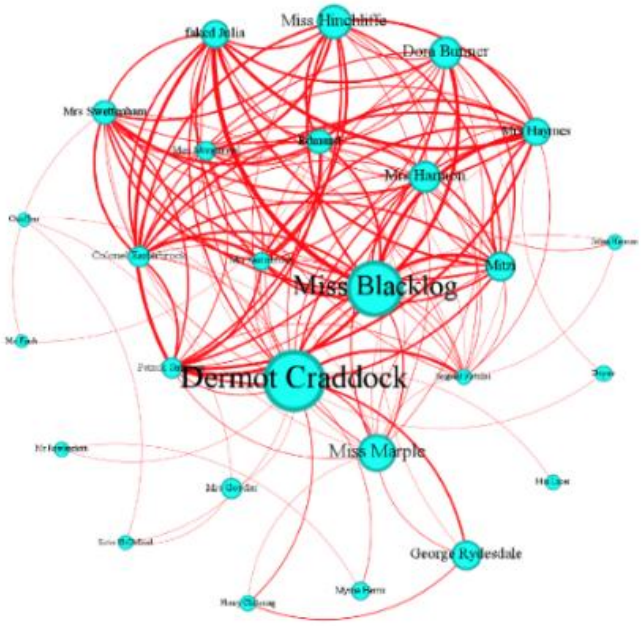

Figure 15: Visualization of Character Relationships in *A Murder is Announced* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

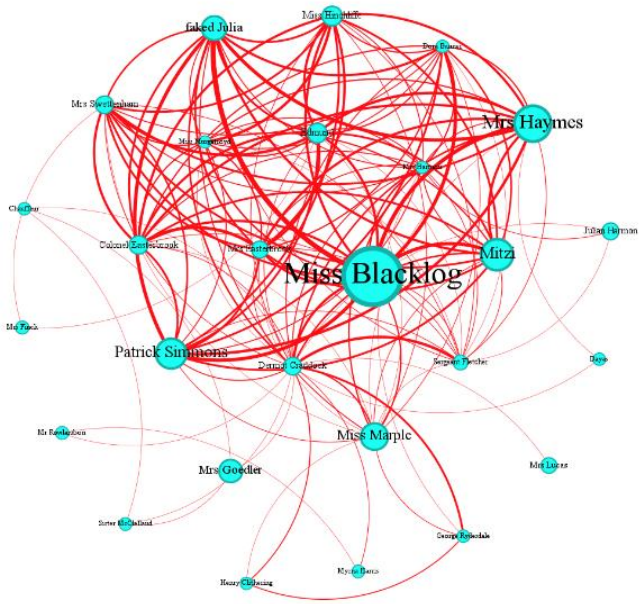

Figure 16: Visualization of Character Relationships in *A Murder is Announced* with Nodes Ranked by Mention, Layout: Yifan Hu

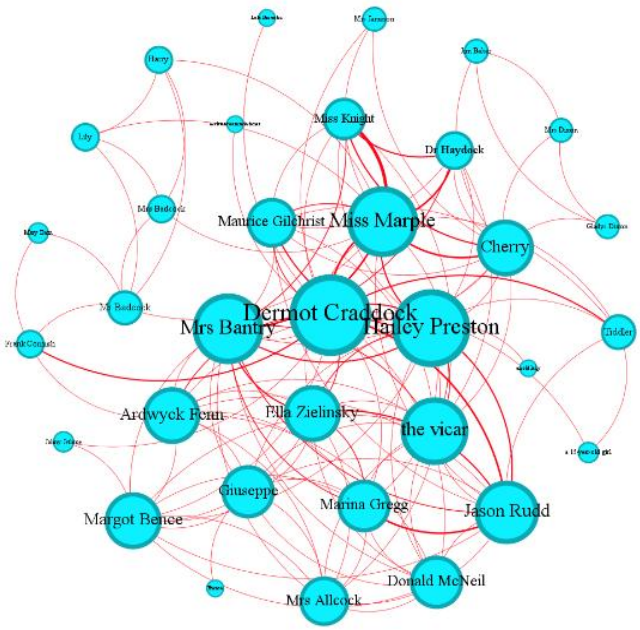

Figure 17: Visualization of Character Relationships in *The Mirror Cracked Side from Side* with Nodes Ranked by Degree, Layout: Yifan Hu

Marina Gregg, the murderer of three victims, does not appear frequently throughout the novel (see Figure 17). After committing her first crime, she pretends to be ill, presenting herself as the intended target of the murder, and is portrayed as extremely frightened and unsettled that she is unable to participate in investigation. Her feigned vulnerability creates the perfect opportunity for her to discreetly eliminate the two individuals: Giuseppe and Miss Zielinksy, who possess compromising information about her and are secretly blackmailing her.

The small node of Marina Gregg, when ranked by Quoted Speeches, is as expected (see Figure 18). She does not appear frequently in the narrative. Thus, the total number of her quoted speeches is not significantly high.

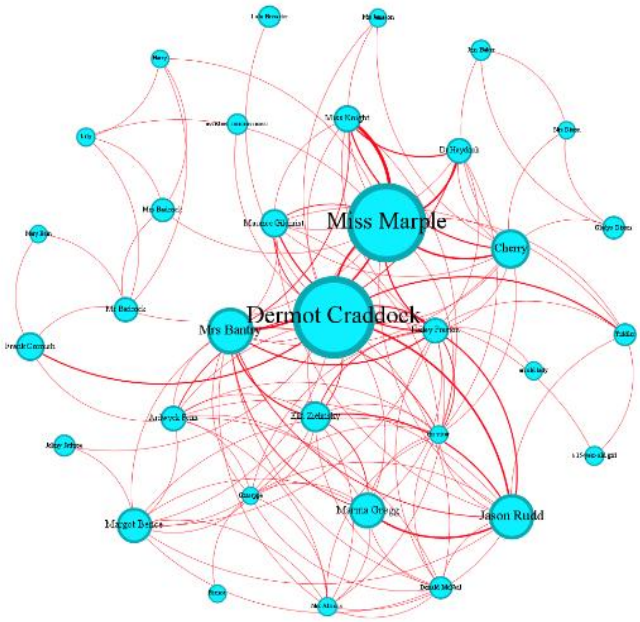

Figure 18: Visualization of Character Relationships in *The Mirror Cracked Side from Side* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

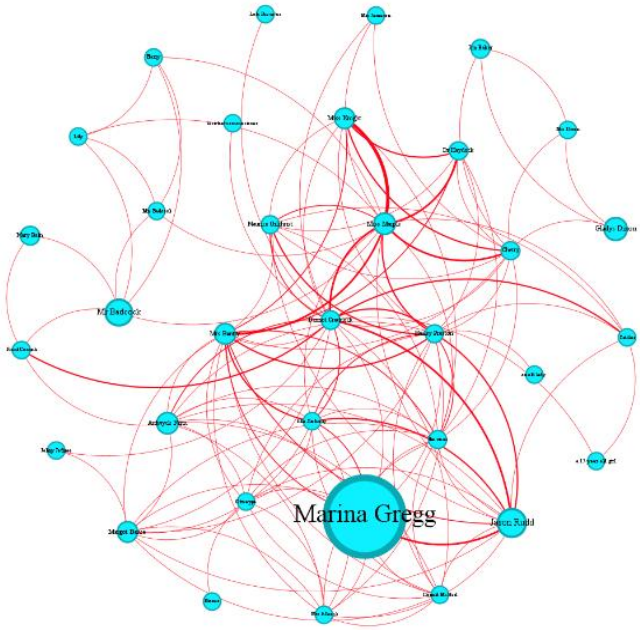

Figure 19: Visualization of Character Relationships in *The Mirror Cracked Side from Side* with Nodes Ranked by Mention, Layout: Yifan Hu

Marina Gregg is the most frequently mentioned character in the novel (see Figure 19). A central question in Miss Marple and Inspector Craddock's investigation is what Marina Gregg was staring at during the cocktail party. Since she is feigning illness due to the staged murder, she is unable to provide much information herself, so the investigators must rely on the accounts of those who attended the party. When it is revealed that she was looking at a painting of a child, her history of having contracted German measles and giving birth to a child with defects comes to light, prompting further investigation into this aspect of her past. Naturally, this line of inquiry results in her being mentioned more frequently than any other character in the narrative. Since Marina Gregg is indeed the murderer, Mention serves again as a validate metric indicating suspect or attention.

Marina Gregg's Closeness Centrality is at a moderate level (0.52), reflecting her close connections primarily with the attendees of the cocktail party. She has no involvement with Miss Marple's social circle. Her Betweenness Centrality is very low (1.93) compared with Inspector Craddock (149.26) and Miss Marple (121.4). As noted previously, Marina Gregg is connected only to the party attendees and does not serve as a bridge for information flow between other characters. Due to her relation to Hailey Preston, who is the butler of Marina Gregg and Jason Rud (Marina's husband) and who has the highest Eigenvector Centrality (1.0), Marina Gregg's Eigenvector Centrality is also at a high level of 0.75.

The Mystery of the Blue Train

The co-murderers in *The Mystery of the Blue Train* are Major Knighton and Ada Mason. In Figure 20, Knighton's node is larger than expected, with a Degree Centrality value of 18. (The largest node has a value of 20) Despite his seemingly quiet demeanor and his role as secretary to Rufus Van Aldin, who is the father of the victim, Ruth Kettering, Knighton's relatively high Degree suggests he is active and well-connected in the narrative. In contrast, although Mason plays a role as Knighton's accomplice, her node is significantly smaller, which reflects her limited overt interactions with others in the story.

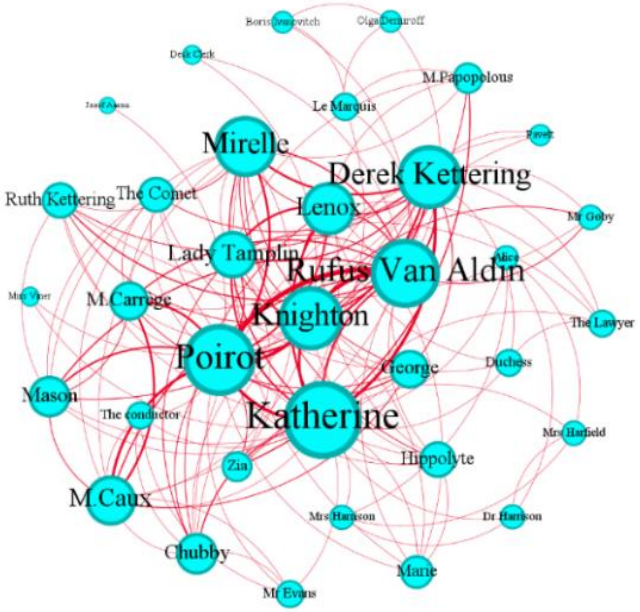

Figure 20: Visualization of Character Relationships in *The Mystery of the Blue Train* with Nodes Ranked by Degree, Layout: Yifan Hu

With respect to Quoted Speeches (see Figure 21), the largest node undoubtedly belongs to Hercule Poirot. As the detective, Poirot is naturally required to engage extensively with other characters, conducting interviews, asking questions, and discussing his deductions, which makes him the most verbally active figure in the narrative. In contrast, the nodes representing Knighton and Mason are comparatively small, suggesting that both characters have limited direct verbal interaction with others. This pattern aligns with their roles as co-murderers, as maintaining a low profile and minimizing communication would be consistent with efforts to avoid suspicion.

In Figure 22, the nodes representing Derek Kettering, the husband of the victim, Ruth Kettering, and The Comet, the secret lover of Ruth Kettering, are among the largest in the network. Both are portrayed as highly suspicious figures: Derek Kettering stands to gain a substantial inheritance upon Ruth's death and has a clear motive to eliminate her to live freely with his lover, Mirelle. Similarly, The Comet is positioned as someone who could have deceived Ruth into handing over her wealth before ultimately killing her.

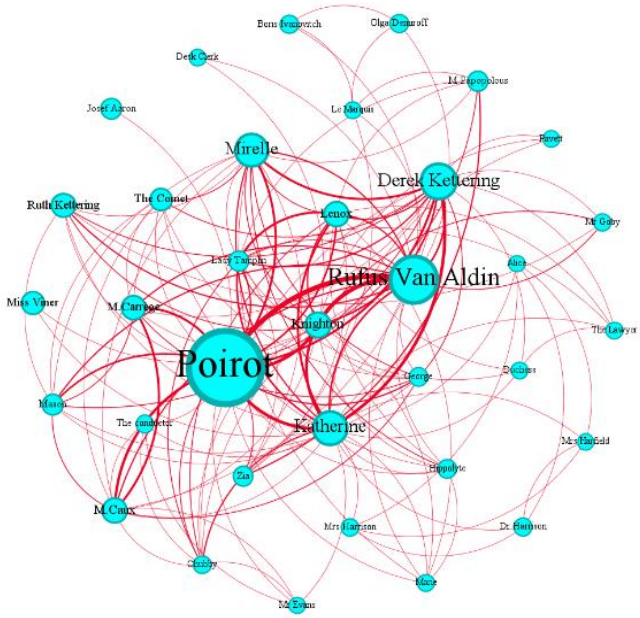

Figure 21: Visualization of Character Relationships in *The Mystery of the Blue Train* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

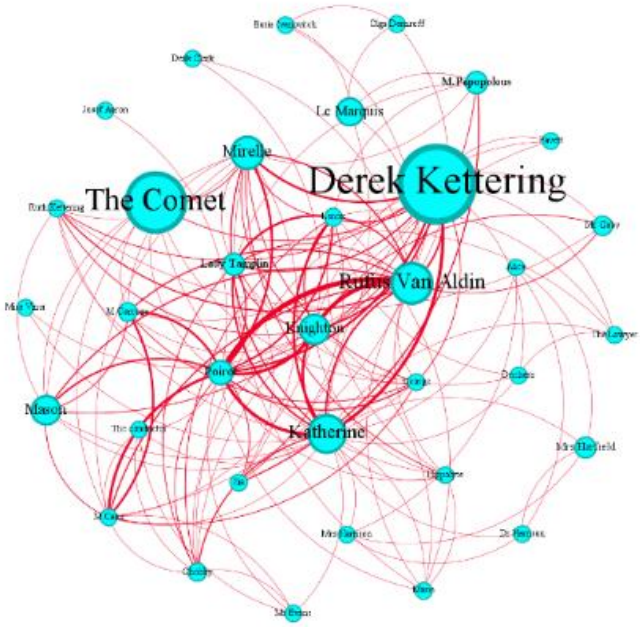

Figure 22: Visualization of Character Relationships in *The Mystery of the Blue Train* with Nodes Ranked by Mention, Layout: Yifan Hu

In contrast, the nodes for Knighton and Mason in this visualization are relatively small. Both characters remain quiet, inconspicuous, and draw little attention from others. These traits contribute to their effectiveness as murderers.

Knighton's Closeness Centrality has an average value of 2, with the highest value being 3, indicating that while Knighton is not particularly close to any other node in the network, he remains relatively well-connected. Additionally, his Betweenness Centrality (47.42) and Eigenvector Centrality (0.88) are elevated compared to other characters, although neither represents the highest values in the network. These centrality measures reflect Knighton's role as secretary to Rufus Van Aldin, where he is tasked with performing duties on behalf of his boss, as well as his connection to Katherine, the character with the highest Eigenvector Centrality (1) and Betweenness Centrality (171).

Mason's Closeness Centrality is identical to Knighton's, indicating that she, too, is not particularly close to anyone, except for her mistress, Ruth Kettering. Furthermore, Mason's Betweenness Centrality and Eigenvector Centrality are both at relatively negligible level due to her sparse dialogues and role as a maid.

In summary, Knighton is more well-connected and active than Mason throughout the novel. In contrast, Ada Mason's involvement primarily occurs after the murder, where she is responsible for concealing the evidence, impersonating Ruth Kettering, and constructing an alibi. Despite their respective actions, both characters remain deliberately unremarkable and inconspicuous in their roles, blending seamlessly into the background of the crime. It is also noteworthy that, although being co-murderers, Knighton and Mason do not co-occur in the network, which highlights a limitation of network analysis of co-occurrence-in-chapter based relationships. As Schumacher (2018) points out, a lack of observable co-action in a network does not equate to the absence of interaction or connection.

Evil under the Sun

In Figure 23, co-murderers, Christine Redfern and Patrick Redfern have the same Degree Centrality value of 16 (with the highest value being 18, attributed to Poirot). Both characters appear with similar frequency throughout the novel, suggesting that they share comparable levels of connectivity within the narrative.

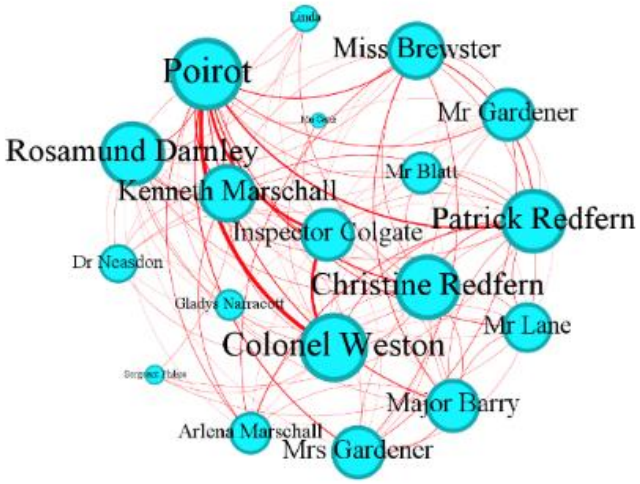

Figure 23: Visualization of Character Relationships in *Evil under the Sun* with Nodes Ranked by Degree, Layout: Yifan Hu

As illustrated in Figure 24, Christine Redfern and Patrick Redfern have considerably fewer spoken lines than Poirot (1009, the highest) and Colonel Weston (647, the second highest), who are actively involved in investigating the case. However, the number of their quoted speeches falls within a relatively moderate range, with Christine Redfern having 229 quotes and Patrick Redfern 224. This suggests that, while they are not as vocal as the primary investigators, both characters maintain a notable presence in the narrative, with similar levels of interaction.

Kenneth Marschall, the victim's husband, and Linda, the victim's stepdaughter, are the most frequently mentioned characters in the narrative (see Figure 25). Kenneth Marschall is portrayed as having an admiration for Rosamund Darnley, while Linda consistently expresses a strong dislike for her stepmother. These position both characters as highly suspicious figures in the story. In addition, the numbers of Mention of Christine Redfern (136) and Patrick Redfern (117) are also noteworthy, suggesting their suspicious roles as co-murderers in the crime.

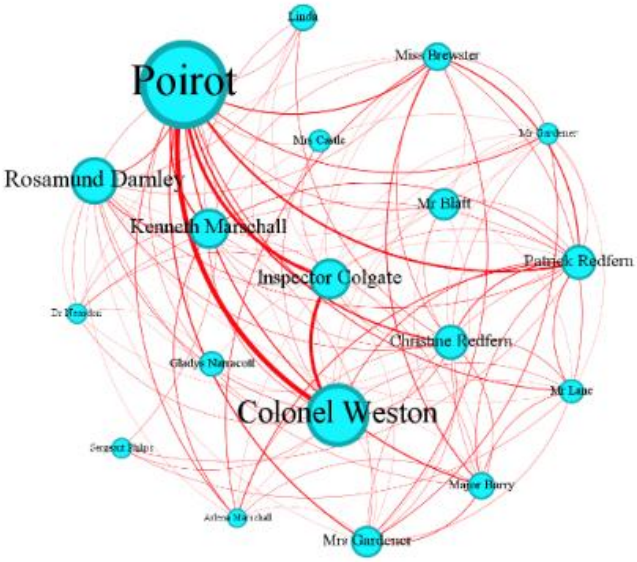

Figure 24: Visualization of Character Relationships in *Evil under the Sun* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

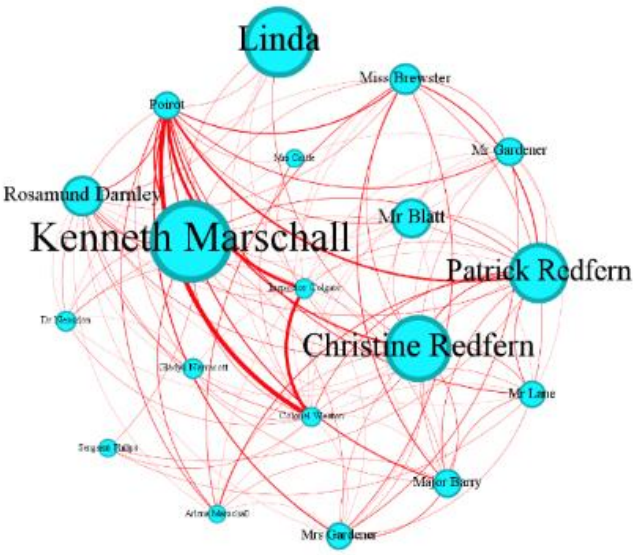

Figure 25: Visualization of Character Relationships in *Evil under the Sun* with Nodes Ranked by Mention, Layout: Yifan Hu

It is noteworthy that the Closeness Centrality of both Christine Redfern and Patrick Redfern is identical (0.94), as are their Eigenvector Centrality values (Christine Redfern: 0.96, Patrick Redfern: 0.95). This similarity is understandable, given that they are a married couple, sharing nearly identical frequencies of appearance, quoted speeches, and mentions throughout the narrative.

Christine Redfern's Betweenness Centrality (4.97) is slightly lower than Patrick Redfern's (6.43). This is due to Patrick's superficial personality, for instance, he pays great attention to his charming appearance to attract women and is highly sociable, using his social interactions to exploit them. By contrast, Christine plays more of an intermediary role, acting as an accomplice who helps conceal Patrick's actions without being directly involved in the more pivotal connections. This aligns with the dynamics of a co-murderer relationship, in which one individual occupies a central, orchestrating position, while the other functions in a supporting capacity.

4.1.4 Synthesis of Results of SNA

In this section, a synthesis of results of SNA is presented, providing comprehensive view of the measurement values of murderers in these eight selected novels. Table 1 presents the metrics for each murderer, with three values provided in each metric column: the murderer's value, the average value for all characters (Ave.), and the highest value among all characters (Max). Due to space limitations, *QS* stands for Quoted Speeches, *Men* stands for Mention, *CC* for Closeness Centrality, *BC* for Betweenness Centrality, and *EC* for Eigenvector Centrality.

Murderer	Gender	Degree/Ave. Degree/Max	QS/Ave.QS /Max	Men/Ave.Men /Max	CC/Ave.CC /Max	BC/Ave.BC /Max	EC/AVE.EC /Max
Sheppard	male	21/11 /21	699/193 /1361(Poi.)	0/49 /285(R. Paton)	1/0.69 /1	48.69/5.23 /48.69	1/0.63 /1
F. Clarke	male	11/8 /37 (P, H)	229/81 /1557(P)	20/11 /178 (Cust)	0.54/0.51 /0.81 (P, H)	4.57/24.63 /319.39 (P, H)	0.47/0.3 /1(P, H)
Wargrave	male	13/10 /14 (Lom.)	349/181 /599 (Lom.)	23/23 /77 (Arms.)	0.96/0.91 /1 (Lomb.)	3.3/2.21 /16.3(Lomb.)	0.99/0.85 /1 (Lomb.)

Jane Wilkinson	female	14/8 /25 (P, H)[5]	272/186 /2047 (P, H)	418/37 /418	0.69/0.62 /1(P, H)	9.4/8.35 /88.83(P, H)	0.75/0.47 /1(P, H)
Miss Blacklog	female	15/11 /25 (Crad.)	647/175 /742 (Crad.)	198/30 /198	0.7/0.63 /0.96 (Crad.)	0.07/8.11 /160.4 (Crad.)	0.92/0.59 /1 (Crad.)
Marina Gregg	female	11/8 /19 (Crad.)	251/143 /884 (Crad.)	628/35 /628	0.52/0.48 /0.69 (Crad.)	1.9/19.6 /149.3 (Crad.)	0.75/0.42 /1.0 (Preston)
Knighton	male	18/8 /23(Kath.)	152/119 /947 (P)	45/23 /209 (Derek)	0.68/0.54 /0.76 (Kath.)	47.42/15.34 /171 (Kath.)	0.88/0.44 /1(P)
Mason	female	10/8 /23 (Kath.)	73/110 /947 (P)	43/23 /209 (Derek)	0.59/0.54 /0.76 (Kath.)	3.59/15.34 /171 (Kath.)	0.61/0.44 /1 (P)
P. Redfern	male	16/12 /18 (P)	224/220 /1009 (P)	117/48 /181 (Kenn.)	0.9/0.76 /1 (P)	4.97/3.1 /15.21 (P)	0.95/0.7 /1 (P)
C. Redfern	female	16/12 /18 (P)	229/220 /1009 (P)	136/48.1 /181 (Kenn.)	0.9/0.76 /1 (P)	4.97/3.1 /15.21 (P)	0.96/0.7 /1 (P)

Table 1: Synthesis of Network Analyses of Murderers' Measurement Values in Selected Novels

Upon examining the data for female murderers (excluding C. Redfern and Mason) in Table 1, it can be observed that their Degree and Quoted Speeches are significantly higher than the average, though not the highest. This suggests that these female murderers are relatively active within the novels. Mention values of these three female murderers are the highest within their respective networks. As written before, Mention is a key indicator of suspicion or attention, as it represents how often the characters are referenced by others within the narrative: both Miss Blacklog and Marina Gregg deliberately stage themselves as the intended targets of the murders, which shifts the investigation's focus toward uncovering who might have the strongest motive to kill them. As a result, they are mentioned frequently throughout the narratives. Similarly, Jane Wilkinson's multiple identities, complex social relationships, and her conflict with the first victim position her as one of the most frequently referenced characters during the investigation. Furthermore, the Closeness Centrality values for these three female murderers are above the average, which implies that they maintain strong connections with other characters. Their Eigenvector Centrality values are also relatively high, indicating that they are connected to resource-rich characters in the novels. For instance, they interact with detectives and inspectors; in

[5] P stands for Poirot, H for Hastings

The Mirror Cracked Side from Side, Preston, who has the highest Eigenvector Centrality value, is Marina Gregg's butler, leading to a high Eigenvector Centrality of Marina Gregg. However, the Betweenness Centrality values for all three female murderers are not high, suggesting that they do not play a significant role in bridging different social groups or transmitting information. Miss Blacklog's social circle is limited to those attending the murder party, and Marina Gregg's is similarly confined to those present at the cocktail party. As a result, both of their Betweenness Centrality values are well below the average. On the other hand, Jane Wilkinson's Betweenness Centrality is slightly above average, which can be attributed to her multiple identities, such as actress, wife, stepmother, lover, and her connections with various social groups.

Regarding the data for male murderers (excluding Knighton and P. Redfern), all three have above-average values for Degree and Quoted Speeches, suggesting that they are prominent and active characters in the novels. In contrast to the female murderers, their Mention values are relatively low, and the characters with the highest Mention values are indeed the ones most suspected through close reading. Their Closeness Centrality values are all above average (with Sheppard's Closeness Centrality score being 1, as he is the narrator), which indicates strong connections to other characters. Except for Sheppard, both Wargrave and Franklin Clarke have low Betweenness Centrality values, implying that they do not serve as bridging characters in the novels. This is consistent with the pattern observed among the female murderers. Similar to the female murderers, these male murderers have high Eigenvector Centrality scores.

Knighton and Partick Redfern's data are compared with their accomplices, Mason and Christine Redfern respectively. In *The Mystery of the Blue Train*, Knighton appears more frequent and active than Mason, as revealed by his higher Degree and Quoted Speeches values. However, both characters score quite lower than Katherine, suggesting that they possess less conspicuous traits, which allow them to effectively avoid suspicion. Their Mention values are almost identical and significantly lower than that of the most suspicious character, Derek Kettering. Both Closeness Centrality and Eigenvector Centrality values of them are above average, suggesting that they are not peripheral in the novel and due to their low value of Betweenness Centrality, they do not play a role in connecting different groups of people or transmitting information.

Christine Redfern and Patrick Redfern are not only co-murderers but also a couple. Due to their relationship, they share similar values in Degree, Quote Speeches, Mention, Closeness Centrality, Betweenness Centrality, and Eigenvector Centrality. All these values are above average, suggesting that they are active, well-connected, and resourceful characters in the novel. In contrast to Mason, Christine Redfern's values for Quoted Speeches and Mention are slightly higher than her male accomplice: Mason and Christine have assisted their male accomplice to cover up their criminal traits but in different ways: Mason primarily through physical actions, such as rolling up the body and impersonating others, and Christine mainly through verbal actions, such as providing falsified testimonies, provoking the daughter of the victim to a suicide attempt through verbal manipulation.

Overall, the female murderers in these eight novels tend to score above average in Degree, Quoted Speeches, and Mention, highlighting their narrative visibility and activeness. Their high Closeness and Eigenvector Centrality values further indicate strong social integration, while their generally low Betweenness Centrality shows that they operate within rather than between groups. The male murderers show a similar pattern in Degree Centrality, Closeness Centrality, Betweenness Centrality (except for Knighton) and Eigenvector Centrality values but differ in Mention: female murderers, especially solo ones, attract more narrative attention and suspicion, whereas male murderers remain less discussed. It is noteworthy that with respective to co-murderers, both female and male murderers have comparable Mention values. Since the female accomplices do not directly involve in the murder, their comparable Mention values suggest that they are not peripheral figures at all within the narrative.

4.1.5 Reflection of SNA

Applying SNA method to analyze character relationships and visualizing the networks offer a clear and intuitive representation of a novel's relational structure, revealing connections that may be overlooked in close reading. For example, in *The Mystery of the Blue Train*, SNA visualization highlights that the murderer, Knighton, is an unexpectedly active character, which is an insight that would be difficult to discern without this approach. By assigning attributes to nodes, SNA can also illuminate a character's role in ways that traditional reading might miss. For instance, setting the *Mention* attribute shows that

characters, despite limited dialogues or appearances, are often referenced, hinting at their suspicious roles. For example, Marina Gregg in *The Mirror Cracked Side by Side* has a relatively low appearance, nevertheless her Mention value is the highest, signaling her as a character of particular attention. Additionally, algorithmic measures like centrality, as computed by SNA tools like Gephi, provide quantitative insights into each character's constellation within the narrative.

One limitation of SNA in this thesis is that it cannot always capture certain types of interactions. For example, in *The Mystery of the Blue Train*, Knighton and Mason, though co-murderers, do not share a direct relationship in the network. Such co-occurrence-based networks have their own constraints: in cases where characters are present but do not interact, such as chapters in which they neither communicate nor recognize one another, or where spatial and temporal shifts occur, no meaningful relationships can be inferred, which can distort centrality measurements. A more nuanced approach, dividing the novel's events by time or location or relationships, could refine co-occurrence analysis, but implementing such a method across eight novels would be time-consuming and impractical for the scope of this thesis. Nevertheless, this approach can be valuable for future, larger-scale projects.

Another crucial methodological reflection involves the effect of narrative perspective on network metrics. In this thesis, the application of SNA across the corpus is sensitive enough to detect fundamental differences in narrative voice without prior coding. *The Murder of Roger Ackroyd* serves as the critical case: the extreme centrality values of Sheppard, who serves both as the narrator and the murderer, are not a distortion of the method but a direct quantitative signature of Agatha Christie's narrative strategy. His artificially inflated centrality values starkly contrast with the networks of other novels, the narration of which is handled by the observer Hastings or a third-person voice. This demonstrates that SNA can not only map character relationships but also help identify the distinct manipulative logic embedded in a first-person culprit narrative, thereby providing empirical support for the "unreadability" of Christie's text. However, this analysis is based on only eight novels, and therefore has limitations. Future research could extend this study by applying SNA to other novels with different narrative structures or perspectives, to assess the broader applicability of this method. For example, analyzing other works with non-linear narratives, unreliable narrators, or multiple first-person perspectives could reveal how SNA might be used to explore more complex narrative techniques.

4.2 Bar Plot of Speech/Appearance Rate

In this section, the frequency of quoted speeches attributed to each murderer across the chapters in which they appear is depicted through bar plots (see Figure 26a and Figure 26b). The ggplot2 package in R (R Core Team 2023) is employed for its flexibility in creating clear, customizable visualizations, allowing for precise control over axes, colors, and plot aesthetics. The bar plot analysis serves as a method to complement network analysis, making it easier to identify patterns in narrative engagement, such as how murderers' presence varies across the novels. Moreover, this approach facilitates comparison between characters, highlights trends in their narrative prominence, and supports quantitative interpretations in a visually accessible manner.

At first glance of the bar plots for solo murderers, there does not appear to be any discernible pattern in their frequency of appearances or speeches. The bar plot of Knighton and Mason in *The Mystery of the Blue Train* still does not reveal any direct relationship between them. Knighton appears more frequently and is more active in the novel than Mason.

Interestingly, in *The Evil Under the Sun*, when Christine and Patrick appear in the same chapter, Christine tends to speak more often than Patrick. This suggests that Christine is actively engaged in providing false verbal testimony to protect Patrick.

Additional features of their narrative presence can be identified by calculating the speech/appearance rate for each character, derived by dividing the total speech count by the number of chapters in which they appear. When talkativeness is classified using distribution quartiles, an examination shows the 75th percentile is 35.75 speeches per chapter, the 50th is 22.5, and the 25th is 16.5. This non-parametric method establishes robust, data-driven categories resistant to outliers. Based on these values, characters are classified as "highly talkative" if their average number of speeches per chapter exceeds the 75%. Characters whose averages fall between the 25% and 75% are categorized as "moderately talkative," while those below the 25% are considered "rarely talkative." The result is presented in Table 2.

Jane Wilkinson, Marina Gregg, and Franklin Clarke have high speech/appearance rates and belong to "highly talkative" murderers. Despite their relatively low number of appearances,

their high speech/appearance rates suggest that they are featured in intense dialogue scenes. These interactions may be manipulative or persuasive, designed to make these characters more memorable or distracting, potentially serving as red herrings or tools for misdirection.

Miss Blacklog, Sheppard, Mason, Patrick Redfern, and Christine Redfern exhibit as "moderately talkative" speakers, indicating their relatively balanced presence. This suggests that they maintain a consistent, yet unobtrusive role in the narrative, blending into the background in a way that allows them to potentially surprise the reader later.

Knighton and Justice Wargrave are "rarely talkative" speakers. Knighton appears more frequently than Mason, however, his speech/appearance rate is lower, indicating that he has a quiet and restrained demeanor across the novels. He seldom expresses his emotion and only speaks when it is necessary, a pattern that aligns well with his role as a secretary. Wargrave has a high appearance level, but he speaks relatively little every time he appears. Wargrave's persona as a former judge suggests that he might prefer to speak in a controlled and authoritative manner, only when necessary to assert his power. Judges typically speak sparingly in the court, using their words with purpose and precision. This minimalistic approach might be part of Wargrave's psychological manipulation, making his words seem more powerful or weighty when he speaks.

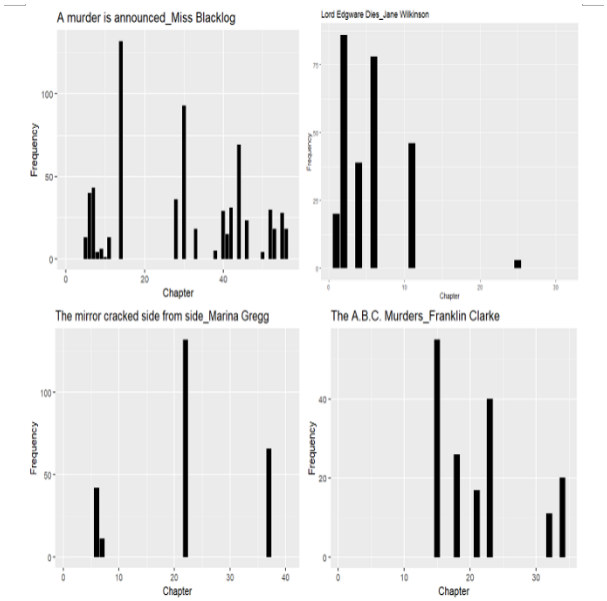

Figure 26a: Bar Plot of Speech/Appearance Rate of Murderers

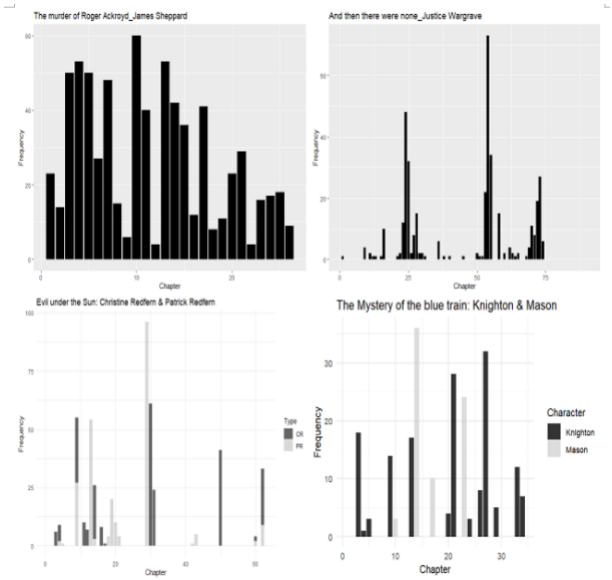

Figure 26b: Bar Plot of Speech/Appearance Rate of Murderers

Murderer	Number of Chapters in which she/he appears/Chapters of Novel	Average number of speeches in each chapter	Level of Talkativeness
Jane Wilkinson	6	45	highly
Miss Blacklog	22	29	moderate
Marina Gregg	4	63	highly
Sheppard	26	27	moderate
F. Clarke	6	38	highly
Wargrave	41	9	rare
Knighton	13	12	rare
Mason	4	18	moderate
P. Redfern	14	16	moderate [6]
C. Redfern	13	18	moderate

Table 2: Murderers' Average Number of Speeches per Chapter of Appearance

4.3 Word2Vec Contextualization

To investigate the semantic fields of the murderers, i.e., how are murderers contextualized across the novels, Word2Vec method is employed. Developed by Mikolov et al. (2013),

[6] Patrick Redfern's rate lies only 0.5% below the 25th percentile, so he is categorized as a moderate speaker

Word2Vec offers an efficient way to calculate word representations in a vector space using artificial neural networks. This method analyzes large text corpora to model the semantic relationships between words and their contexts, and is widely used in text analysis and natural language processing (NLP).

The method is based on the distributional hypothesis, which suggests that words with similar meanings tend to appear in similar contexts (Firth, 1957; Bojanowski et al. 2017; Gius et al. 2021). To visualize and analyze these relationships, dimensionality reduction techniques like Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (t-SNE) are often applied (Schumacher, 2023a).

Word2Vec employs two primary models to learn these word representations: the Continuous Bag-of-Words (CBOW) model and the Skip-Gram model. In CBOW, the algorithm predicts a target word based on its surrounding words, while in Skip-Gram, it predicts the surrounding words based on a given target word. Both methods help map words into a multi-dimensional space where words with similar meanings are placed closer together, based on shared context.

4.3.1 Corpus Preparation and Model Training

Since Word2Vec is typically applied to large corpora, using it on individual novels in this thesis may produce unreliable results. To address issues arising from the limited size of each novel, the eight selected works are combined into a single dataset. This approach increases the overall data volume, enabling the detection of broader patterns in Agatha Christie's characterizations of murderers across multiple novels. For example, it allows us to examine whether murderers share traits with other characters or are consistently portrayed as distinctive, shedding light on whether Agatha Christie intentionally individualizes her murderers or blends them with other characters to enhance their unpredictability and narrative complexity.

In this study, contextualization of non-murderers is thus represented as comparison. Non-murderers in this study refer to detectives and the most suspicious characters (who are with the highest score of Mention) in each novel. Since Miss Blacklog, Jane Wilkinson, and Marina Gregg are already identified as both murderers and most frequently mentioned

characters within their respective narratives, the second most suspicious characters are chosen from these narratives: Phillipa from *A Murder is Announced*, Ronald Marsh from *Lord Edgware Dies*, and Jason Rudd from *The Mirror Cracked Side from Side*. The names of all target characters in Word2Vec analysis are standardized to avoid identical names. For example, since Miss Marple and Jane Wilkinson share the same first name *Jane*, all references and forms of address referring specifically to Jane Wilkinson are normalized to *JaneWilkinson* to ensure consistency.

To prepare the corpus for modeling semantic relations, tokenization is first performed at the sentence level to segment the text. Subsequently, word-level tokenization is applied to ensure that the model can process individual token within their respective contexts. A Word2Vec vector model is then trained using the Gensim package. This procedure follows methodological guidelines outlined in the *forTEXT* learning unit *Word2Vec with Gensim* (Schumacher, 2023b). The model learns distributed representations of words by analyzing their co-occurrence patterns across the corpus.

The trained vector model is used to identify words that occur in similar contexts to each target character. Using targeted queries (via the *most_similar()* function in Gensim), the words with the highest similarity values to each character are retrieved. Figure 27 illustrates the top 20 most similar words to the murderer, Miss Blacklog, as an example, with each word accompanied by its similarity value.

```
[('Russell', 0.9004054665565491),
 ('Ganett', 0.892592191696167),
 ('Marple', 0.890663743019104),
 ('Viner', 0.881113588809967),
 ('Adams', 0.8697505593299866),
 ('Grey', 0.868432343006134),
 ('Murgatroyd', 0.8632226586341858),
 ('Bunner', 0.8622099161148071),
 ('Knight', 0.8558052182197571),
 ('Hinchliffe', 0.8485541343688965),
 ('MarinaGregg', 0.8379019498825073),
 ('Carroll', 0.817135214805603),
 ('Ackroyd', 0.8153860569000244),
 ('Merrion', 0.806546151638031),
 ('Brady', 0.8063899278640747),
 ('Brewster', 0.797980785369873),
 ('Claythorne', 0.7944421768188477),
 ('Darnley', 0.76462322473526),
 ('Oliver', 0.7440248727798462),
 ('Zielinsky', 0.7395014762878418)]
```

Figure 27: Semantic Filed of Miss Blacklog with Top 20 Most Similar Words

4.3.2 t-SNE Visualization

To visualize the relational proximity between words and thereby make patterns of characterizations visible, the semantic fields are represented using t-SNE. Figure 28 is a t-SNE representation of the semantic fields associated with murderers and non-murderers individually, with each character displayed in a different color. In Figure 28, several subtle patterns can be observed. For instance, at the very top center, *Marple*, *MarinaGregg*, *Blacklog*, and *JaneWilkinson* appear in close proximity. The overlap between Miss Marple's semantic field and those of the other characters may be attributed to their shared background: Miss Marple is the amateur sleuth in both Miss Blacklog's and Miss Gregg's cases. The semantic fields of Marina Gregg, Miss Blacklog, and Jane Wilkinson show notable similarities: each is a central figure in her respective novel, and all three mislead the investigation by performing weakness or victimhood. In the above-middle area lies a dense green and yellow mixed color cluster representing mostly *JamesSheppard*. As both the narrator and an apparently neutral figure, Sheppard conceals his guilt so effectively that readers only discover he is the murderer at the end. *JusticeWargrave*'s points (purple colored) are more scattered, often appearing near words such as *commissioner* or *constable*, which aligns with his self-presentation as a symbol of justice in the novel.

Notably, the relationship between Major Knighton and Ada Mason, which cannot be fully captured in SNA, becomes visible in the t-SNE space through their spatial proximity. Likewise, Christine Redfern and Patrick Redfern occupy nearly the same area, reflecting both their marriage and their collaboration as co-murderers.

Figure 29 shows a t-SNE representation of the semantic fields of murderers and non-murderers in groups, with red representing murderers and blue representing non-murderers. It also displays the top 20 most similar words for each character.

Through 29, it is even more obvious that every character: detective, suspect, or murderer, is textually encoded through similar language, tone, and social cues. There is no clear linguistic boundary distinguishing a murderer from anyone else.

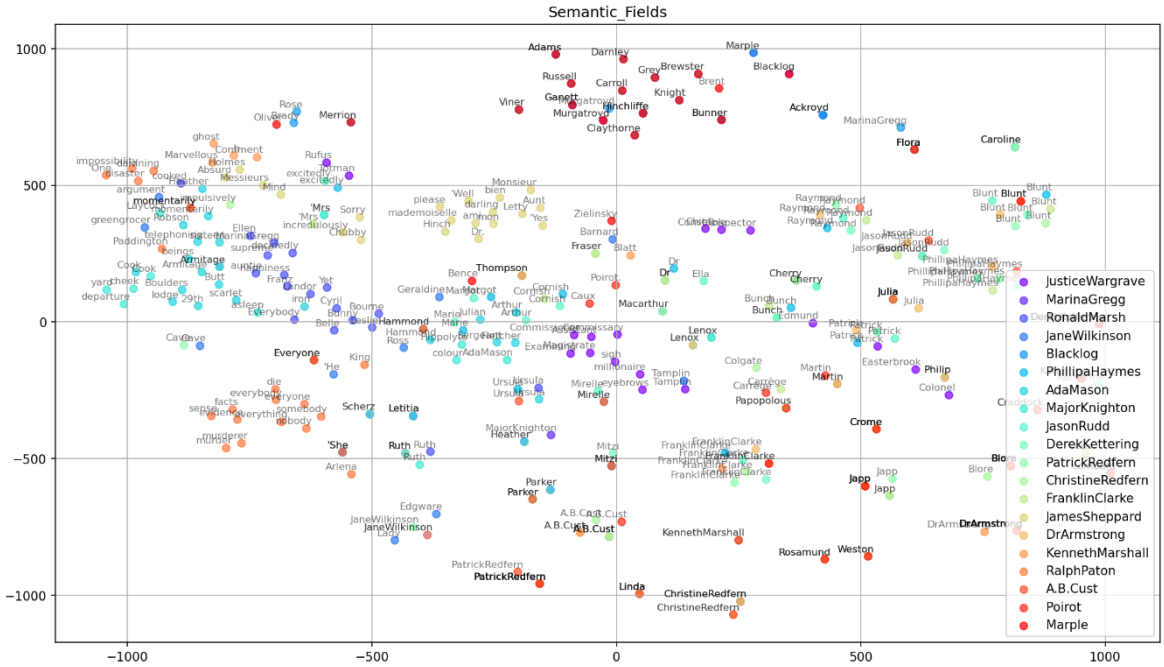

Figure 28: t-SNE Representation of Semantic Fields of Murders and Non-Murderers individually

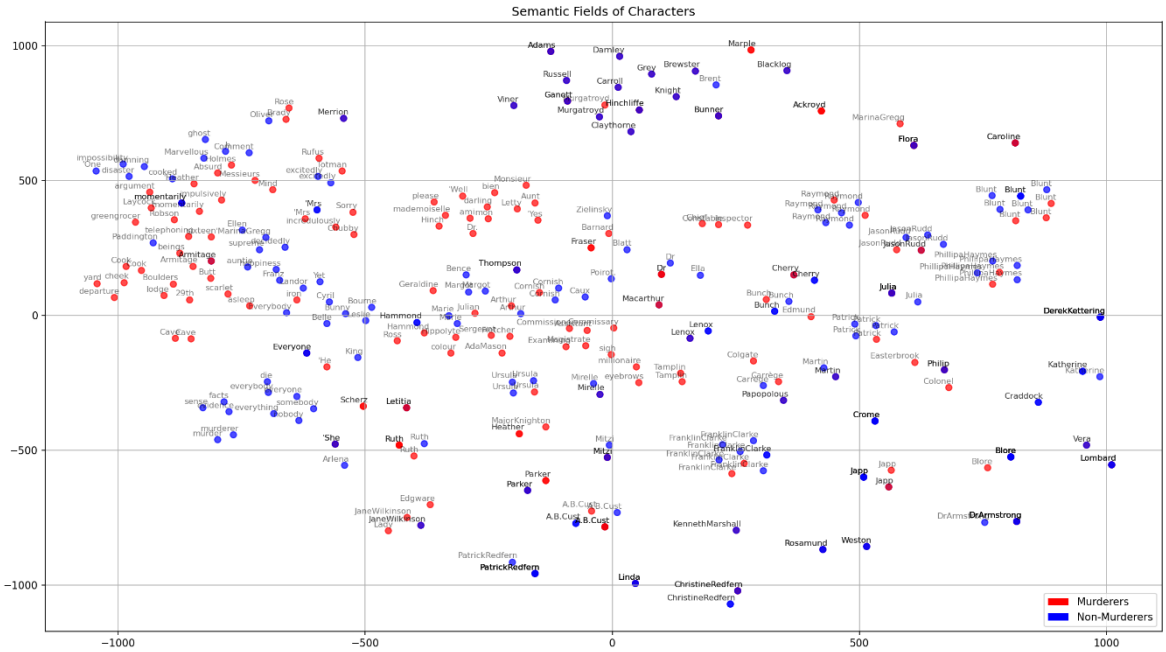

Figure 29: t-SNE Representation of Semantic Fields of Murders and Non-Murderers in Group

Additionally, in this section, a statistic method, i.e., an independent samples t-test, is performed to further explain that there are no distinct boundaries between murderers and

non-murderers. T-test is used to determine whether the means of two groups are significantly different from each other relative to the variability within each group in statistics. This test is appropriate in this section, because the data consists of two independent groups (murderers and non-murderers) and a numerical measure for each character (the mean of the Word2Vec embeddings), allowing a comparison of group means. As the p-value is 0.426, which is considerably higher than the conventional significance level of 0.05, indicating that murderers do not differ statistically from non-murderers in terms of the semantic content of their language.

Overall, Agatha Christie renders her murderers linguistically ordinary: murderers speak and behave like any other character in her fictional world. This supports Ina Hark's qualitative theory of unreadability of Agatha Christie's texts. In data terms, the overlapping embeddings of murderers and non-murderers provide a quantitative reflection of this unreadability.

4.3.3 Reflection of Word2Vec

In applying the Word2Vec model, this study encounters two major challenges. First, as previously mentioned, Word2Vec requires a large volume of texts for effective training. To mitigate the limited size of each individual novel, eight works are combined, resulting in a total of 805,967 tokens. This approach allows for broader cross-novel comparisons and the detection of semantic patterns across Agatha Christie's murderers. However, the total number of tokens is still not big enough to get a more reliable and robust result.

Second, the diversity of contextualized items poses a limitation. Each novel features a different murderer, which constrains the model's ability to learn fully stable and consistent embeddings for individual characters. The embeddings of specific murderers remain sensitive to the unique context of each story.

Nevertheless, Word2Vec still provides valuable insights. For instance, it highlights similarities among the three female murderers who act alone and uncovers relational patterns between characters that are not visible through network analysis. Furthermore, since in word embeddings each word is represented in the form of a numerical vector, it is possible to apply statistical methods to these vectors to quantify linguistic similarities and

differences between groups of characters. Although Word2Vec does not uncover a distinctive pattern among the murderers in this study, this absence of strong clustering paradoxically reinforces the theoretical frameworks of the thesis: the unreadability of Agatha Christie's texts and her continual inventive variation in character construction.

4.4 Sentiment Analysis

The preceding sections examine the relationships that murderers maintain with other characters, how intensive do the murderers engage in the dialogues, and how murderers and non-murderers are contextualized across the novels. Through metric values through SNA, the speech/appearance rates shown in bar plots, an examination of the semantic fields of murderers and non-murderers, Agatha Christie's underlying design principle can be revealed: her murderers are diverse in characterizations, and there is no distinct boundary that separates them from non-murderers. The present section reflects her design principle through sentiment analysis.

4.4.1 Research Tradition

Sentiment analysis is a method used to evaluate the emotional tone of a text by analyzing the words it contains. The technique originated in the 1990s, when researchers began converting written product reviews into numerical rating scores. A common output of sentiment analysis is the Sentiment Polarity Index (SPI), which reflects how positive or negative a piece of text is. This index is typically generated by counting the number of words associated with known positive or negative emotions. (Min & Park, 2019)

Sentiment analysis is also a classic methodology in the scholarship of Digital Humanities. Nalisnick & Baird (2013) introduce an automated method for examining sentiment dynamics among characters in Shakespeare's plays. Leveraging the structured format of dramatic dialogue, they are able to infer with reasonable accuracy which characters are interacting in a given exchange. By determining the likely recipient of each utterance, sentiment can be attributed accordingly. This allows them to identify patterns of alliance and antagonism between characters, as well as to isolate scenes that are pivotal to a

character's emotional trajectory. Cahyadi & Arianto (2024) conduct a computational study that applies sentiment analysis to the characters in *Romeo and Juliet* using the Naïve Bayes algorithm. Their goal is to examine how the emotions expressed by different characters contribute to the development of the play's tragic narrative. They note that, despite the play's emotional richness, no previous research has systematically analyzed character sentiments in this text. To fill this gap, they use a probabilistic Naïve Bayes model to classify the emotional tone (such as positive, neutral, or negative) of textual data associated with each character. Their model achieves 81% accuracy, with relatively strong precision, recall, and F1-scores, particularly for neutral and positive categories. Overall, their study demonstrates how computational methods can reveal patterns in character emotions and enhance the understanding of emotional dynamics in classic literature. Tamanna (2025) uses a mixed-methods approach, combining computational sentiment analysis with historical literary criticism, to explore the relationship between emotional expression and the socio-cultural context in Victorian literature. The focus is on melancholy, a prominent theme in the period, and how it is represented in Victorian texts. This study applies modern text-mining techniques like sentiment lexicons, tokenization, and normalization to track important historical events and socio-political changes in the literature.

Building on this growing body of research, this thesis applies lexicon-based computational sentiment analysis to the linguistic modifiers surrounding the murderers in the selected novels. Unlike previous researches that have focused on the emotional structure of narratives or character interaction patterns, this thesis investigates how sentiment is linguistically constructed around the murderers. The aim is to offer further insights into Agatha Christie's conception of murderers.

4.4.2 Identification of Modifiers

In this thesis, modifiers refer to the lexical items that describe each murderer, including nouns, adjectives, verbs, and associated adverbs. For this step, AntConc (Anthony, 2023) is utilized. AntConc is a powerful tool for corpus linguistics and text analysis. To identify as many modifiers as possible for each murderer, the *Cluster* function is employed in this study. In the context of AntConc, *Cluster* refers to sequences of words that frequently appear together within a corpus.

Before conducting the analysis in AntConc, one critical issue must be addressed: the co-references of the murderers. For example, in *A Murder is Announced*, the murderer Miss Blacklog, is referred to in various ways: *Letti*, *Letty, Miss Blacklog*, *Letitia Blacklog*, and *Charlotte*[7]. To ensure that modifiers of Miss Blacklog can be accurately identified as many as possible, it is essential to standardize all references to a single name. In this study, all co-references of Miss Blacklog are replaced with *Blacklog.* Then, the standardized name *Blacklog* is entered into AntConc to generate its word cluster. While explicit references to the characters are standardized through co-reference resolution, pronouns referring to the character (e.g. *she*, *her*) are not fully resolved due to the limitations of automated coreference systems and the scale of the corpus. Consequently, descriptions or modifiers linked to the characters may be underrepresented in some passages. This limitation, however, affects all characters equally and does not undermine the comparative trends observed. An example of the output of AntConc is represented in Figure 30.

	Cluster	Rank	Freq	Range
1	miss blacklog	1	309	1
2	blacklog s	2	30	1
3	blacklog had	3	26	1
4	blacklog i	4	21	1
5	blacklog and	5	18	1
6	blacklog was	6	14	1
7	blacklog but	7	12	1
8	blacklog said	7	12	1
9	blacklog she	7	12	1
10	blacklog with	7	12	1

Cluster	Rank	Freq	Range
poor blacklog	84	1	1
room blacklog	84	1	1
said blacklog	84	1	1
see blacklog	84	1	1
shop blacklog	84	1	1
shoulders blacklog	84	1	1
soon blacklog	84	1	1
stairs blacklog	84	1	1
state blacklog	84	1	1
the blacklog	84	1	1

Figure 30: A Screenshot of Cluster for the Murderer Miss Blacklog in AntConc

Figure 30 presents part of the result of a search conducted using the *Cluster* function in AntConc. The search range is set to include one word before and after the term *Blacklog*, resulting in 236 hits in total. It is important to note that not all these 236 word-combinations are modifiers of *Blacklog*, and some potential modifiers may still be missing. To address this issue, the output must be manually reviewed and refined. To enhance the

[7]The murderer, Miss Blacklog, is actually Charlotte Blacklog, who is disguised as her sister, Letitia Blacklog. However, whenever the sister is referred to, only *Letitia* is used. Therefore, *Letitia* alone is kept as it is, only *Letitia Blacklog* and all other co-references should be standardized into *Blacklog*.

efficiency of manual control, certain parts of speech, such as prepositions (e.g., *at, in, on*), interjections (e.g., *oh*), pronouns (e.g., *he, I*), adverbs (e.g., *yes, no, but*), and conjunctions (e.g., *and*), can be excluded first. Next, if the word *Blacklog* is followed by words like verbs, it is necessary to further investigate the context of these combinations using the KWIC (Keyword in Context) function in AntConc. As shown in Figure 31, several adverbs that follow *said* and function as modifiers of *Blacklog* are identified, such as *slowly, anxiously, grimly, quickly*.

Left Context	Hit	Right Context
seen was the blinding light of the electric torch." Miss	Blacklog said	slowly, "And you believe that one of those
idea, anyway?" said Julia, yawning. "What does it mean?" Miss	Blacklog said	slowly, "I suppose - it's some silly sort
just seems gibberish to me." Craddock replaced the receiver. Miss	Blacklog said	anxiously: "Has something happened to Miss Marple? Oh,
Marple took the note from Bunch's hand. "Tell Miss	Blacklog," said	Bunch, "that Julian is terribly sorry he can'
is from your sister Julia?" "Yes - yes, it is." Miss	Blacklog said	grimly: "Then who, may I ask, is the
a side of bacon." "Show him that note from Miss	Blacklog," said	Miss Marple. "It's some time ago now,
I'm nothing but a trial to you, Letty." Miss	Blacklog said	quickly, "You're my great comfort, Dora. And
Who gets your money if you were to die?" Miss	Blacklog said	rather reluctantly: "Patrick and Julia. I've left
word 'Liar.' As though she had read his mind Miss	Blacklog said: "	Please don't be too prejudiced against the
Somebody's murdering Mitzi..." Patrick muttered: "No such luck." Miss	Blacklog said: "	We must get candles. Patrick, will you -" The
Patrick affectionately. "I'll look after you." "You?" was all	Blacklog said,	but the disillusionment behind the word was almost
what she looked like?" "She was small and dark, Miss	Blacklog said." "	Really," said Miss Marple, "that's very interesting." "

Figure 31: A Screenshot of KWIC Result for the Word Sequence *Blacklog said* in AntConc

The same process is applied to all the other murderers, with the search results being manually recorded. To facilitate a comparison between the detectives and the murderers-representing good and evil respectively-in the subsequent sentiment analysis, the modifiers of both Poirot and Miss Marple are also examined.

For the purpose of sentiment analysis and the construction of sentiment networks, for each character: murderers and detectives, a CSV file is created, with character name as *target* and its modifiers as *source*. In cases where verbs are modified by adverbs, the verbs are first treated as *source*, then as t*arget* with the adverbs as the *source*. An example is represented in Figure 32.

target,type,source
Blacklog,undirected,darling
Blacklog,undirected,dearest
Blacklog,undirected,poor
Blacklog,undirected,wear
wear,undirected, always pearl
Blacklog,undirected,became
became,undirected,uneasy
Blacklog,undirected,began to understand
Blacklog,undirected,came
came,undirected,out
came,undirected,in
comes,undirected,along
Blacklog,undirected,bought
bought,undirected,heavy glass
Blacklog,undirected,cared
care,undirected,about murders
Blacklog,undirected,annoyed
Blacklog,undirected,said
said,undirected,cheerfully

Figure 32: A Screenshot of *Blacklog*'s Modifiers in CSV Format

4.4.3 Result of Sentiment Analysis

Sentiment analysis in this section aims to explore Agatha Christie's characterization of her murderers by classifying the tone of the modifiers as negative, positive, or neutral. This analysis is conducted using a dictionary-based automatic classification method implemented in R (R Core Team 2020). The Lexicoder Sentiment Dictionary (LSD) developed by Young & Soroka is employed for this purpose. LSD contains 2,858 negative and 1,709 positive sentiment words, along with 2,860 (1,721) negations of negative (positive) words, allowing it to capture nuanced sentiment shifts. (Young & Soroka 2012)

While recent approaches such as large language models (LLMs) offer high flexibility and context sensitivity, LSD provides several advantages for the present study. First, it ensures transparency, as every word included in the dictionary and its associated sentiment category are explicitly defined. This allows readers and future researchers to trace the coding decisions, thereby enhancing replicability. Second, the LSD has been empirically validated against human-coded news content and shown to produce results that align more systematically with human judgment than many other sentiment dictionaries (Young & Soroka, 2012), providing confidence in its reliability. Third, the LSD recognizes inflected and declined forms, meaning that modifiers do not need to be lemmatized during preprocessing. This reduces the risk of data loss during normalization and simplifies the subsequent automated analysis. Therefore, although LLMs could, in principle, offer richer, context-sensitive sentiment interpretation, the LSD is chosen for its transparency,

replicability, and demonstrated reliability, making it particularly well-suited to this study. A screenshot of the result of the sentiment analysis for the murderer Miss Blacklog is represented below as an example (see Figure 33).

To generate a sentiment network, the CSV files created for the modifiers of each target character are first imported into Gephi (Bastian et al. 2009). In Gephi, *sentiment* is then added as an attribute in Gephi labor, and the sentiment scores are then as values imported. For a clear division of sentiment, the score of negative sentiment is changed to -1, positive sentiment remains as 1, and neutral sentiment is represented by a score of 0.

```
source,"negative","positive","neg_positive","neg_negative"
darling,0,1,0,0
dearest,0,1,0,0
poor,1,0,0,0
wear ,0,0,0,0
always pearl,0,0,0,0
became ,0,0,0,0
uneasy,1,0,0,0
began to understand,0,1,0,0
came,0,0,0,0
out,0,0,0,0
in,0,0,0,0
along,0,0,0,0
bought ,0,0,0,0
heavy glass,1,0,0,0
cared,0,1,0,0
about murders,1,0,0,0
annoyed,1,0,0,0
```

Figure 33: A Screenshot of Sentiment Analysis for the Modifiers of Miss Blacklog

In this thesis, while different sizes of nodes in SNA are important for interpretation, the total sentiment scores of the target character in each sentiment network is of greater importance, so not all visualizations of sentiment networks are represented. As an example, visualizations of sentiment networks for Miss Blacklog and Miss Marple in *A Murder is Announced* are selected. All visualizations can be found in GitHub repository.

The nodes in the sentiment networks represent the modifiers of the corresponding characters and are color-coded according to their sentiment scores. In these visualizations, purple indicates neutral sentiment, red indicates positive sentiment, and green indicates negative sentiment (see Figure 34-35). For Miss Blacklog, positive nodes account for 25.76% of the total, negative nodes for 21.21%, and neutral nodes for 53.03%. As a result,

Miss Blacklog has a sentiment score of 4.55[8], suggesting that Agatha Christie portrays her in a predominantly positive light. For Miss Marple, positive nodes make up 17.57% of the total, negative nodes 9.46%, and neutral nodes 72.97% and her sentiment score is 8.11, unequivocally suggesting that Agatha Christie treats Miss Marple as a positive character.

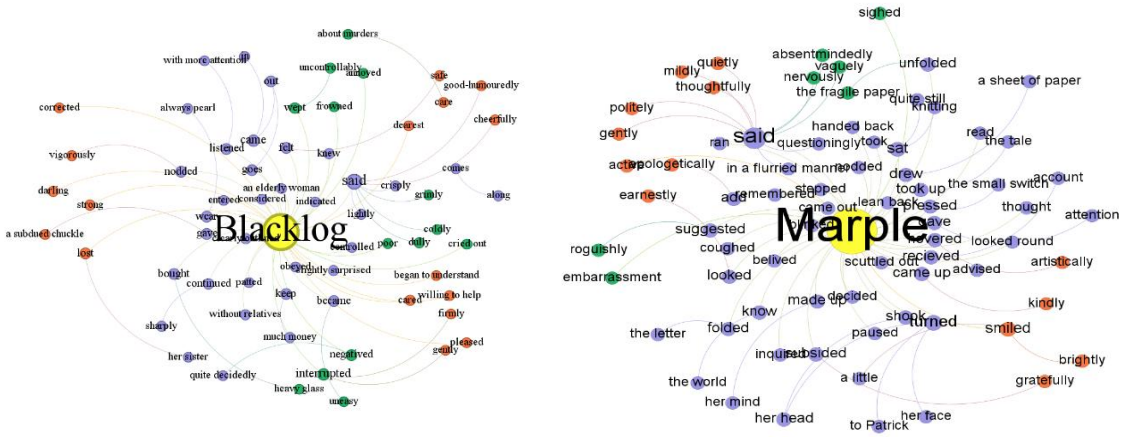

Figure 34-35: Sentiment Network Visualizations for Miss Blacklog and Miss Marple in Gephi, Layout: Yifan Hu

The result of sentiment analysis of all murderers and detectives are summarized in Table 3.

	Neutral (%)	Positive (%)	Negative (%)	Total Score
The Murder of Roger Ackroy				
Sheppard	40	30	30	0
Poirot	63	26	11	15
The A.B.C. Murders				
F. Clarke	68.89	13.33	17.78	-4.45
Poirot	77.78	14.29	7.94	6.35
And Then There were None				
Wargrave	80.43	11.96	7.61	4.35
Lord Edgware Dies				
Jane Wilkinson	70.59	8.82	20.59	-11.77
Poirot	61.33	26.67	12	14.67
A Murder is Announced				
Miss Blacklog	53.03	25.76	21.21	4.55
Miss Marple	72.97	17.57	9.46	8.11
The Mirror Cracked Side from Side				
Marina Gregg	59.18	22.45	18.37	4.08
Miss Marple	60	24	16	8

[8] The formula for calculating the sentiment score is: neutral * 0 + positive * 1 + negative * (-1).

The Mystery of the Blue Train				
Knighton	54.95	29.67	15.88	13.79
Mason	68.75	9.38	21.88	-12.5
Poirot	64.21	27.37	8.42	18.95
Evil under the Sun				
C. Redfern	76.67	5	18.33	-13.33
P. Redfern	60.78	21.57	17.65	3.92
Poirot	67.59	23.15	9.26	13.89

Table 3: Sentiment Scores of Murderers and Detectives in Selected Novels

It is within expectation that the sentiment scores of Poirot and Miss Marple are far above zero. Detectives in crime novels symbolize goodness and justice, with the task of uncovering and stopping crimes.

In contrast, the murderers exhibit a range of sentiment scores. When combined with close reading, these results become understandable. Sheppard is the only character with a neutral sentiment score. This can be attributed to the fact that Sheppard is the narrator: he deliberately maintains a tone that is neutral and conceals his true emotions, so that when readers eventually discover he is the murderer, they are genuinely surprised. Additionally, throughout the novel, very few modifiers are used to describe him.

Wargrave, Miss Blacklog, Marina Gregg, Knighton, and Patrick all have sentiment scores above zero, which aligns with expectation. Wargrave, a retired judge, represents justice; Miss Blacklog, a wealthy landlady, and Marina, a beautiful and popular actress, both win sympathy and goodwill by pretending to be targets rather than culprits. Knighton's modest behavior and diligent attitude earn him a positive impression; Patrick, outwardly charming, athletic, and sociable, also receives a positive sentiment score.

By contrast, Franklin Clarke, Jane Wilkinson, Mason, and Christine have negative sentiment scores. Agatha Christie provides little description of Franklin, but once he is revealed as the murderer, his revealed attitude and demeanor demonstrate his brutality. Jane Wilkinson is portrayed as a foolish actress; Mason's pretended timidity and helplessness contribute to her negative sentiment score; and Christine is often described by her husband as *ridiculous* and pretends to be hurt by her husband's alleged affair with the victim, which together result in an overall negative sentiment.

These findings lend empirical support to Verbogen (2008)'s argument that Agatha Chritie transcends simple repetition through inventive variations: the murderers can be portrayed not only as negative, but also as positive and neutral. And precisely these variations of

conception of characters reflect Ina Hark's theory of the unreadability of Agatha Christie's texts, which posits that "human guilt and innocence are thoroughly intermixed, rather than being in mutually exclusive binary positions. If we attempt to read people like books to separate the murderous from the non-murderous, we must fail, because in Christie's view, no such creature as a person incapable of murder exists" (Hark 1997, 2).

4.4.4 Reflection of Sentiment Analysis

Sentiment analysis is a valuable method, offering an overview of the sentimental or emotional tone embedded in text. In literary analysis, sentiment analysis can be employed to examine the emotional trajectories and sentiments of central characters, shedding light on the author's conceptualization of the characters as well as their representation within the work. In this thesis, sentiment analysis has yielded meaningful results. The analysis reveals that murder, in Agatha Christie's novels, can be committed by anyone. For instance, in *A Murder is Announced* and *The Mirror Cracked from Side to Side*, both Miss Blacklog and Marina Gregg have committed three murders each. Despite this, their sentiment scores are positive, remaining above zero. This finding can be attributed to their deliberate pretense of being victims in these narratives. They both utilize their established personas, such as being female, *frail*, *sweet*, and friendly, to divert attention and, consequently, quietly achieve their murderous goals. Hence, sentiment analysis, like other methodologies in Digital Humanities, serves as a useful supporting tool for interpreting the characters in literary works.

The effectiveness of a methodology depends on an awareness of its limitations. Sentiment analysis relies on an automatic process grounded in pre-designed sentiment dictionaries, although it eliminates the subjective biases in human review, however, this automated approach has limitations. It is incapable of capturing sentiments that require close reading to discern, such as irony, sarcasm, or context-dependent emotions. Since manually verifying every word after the automated process is impractical, a manual check through sampling can be considered, particularly when dealing with a very large corpus.

Another limitation is that the result of sentiment analysis is influenced by the sentiment dictionary employed. Language is dynamic, constantly evolving, with new words emerging

and old words acquiring new meanings. As such, the continuous updating and refinement of sentiment dictionaries are essential to maintaining the accuracy and relevance of sentiment analysis.

4.5 Linguistic Representation

In this section, linguistic representation of the characters (murderers and detectives) is presented. This representation is based on the modifiers which are also used for the sentiment analysis.

4.5.1 Research Tradition

Linguistic representation refers to the way linguistic elements, such as words, phrases, certain syntax, are used to represent ideas, concepts, experiences, and social realities, thereby shaping how we understand and interpret the world.

Linguistic representation as a methodology is widely applied in Digital Humanities. Burrows (1987) pioneers combination of literary analysis with computational methods to study character linguistically in *Pride and Prejudice*. By examining word frequencies in dialogue, Burrows demonstrates that Austen's characters can be statistically distinguished by their linguistic styles: words like prepositions, pronouns, and conjunctions reveal distinct and meaningful patterns in the speech of individual characters. Newman et. al research "gender differences in language use…using standardized categories to analyze a database of over 14,000 text files from 70 separate studies." (Newman et al. 2008, abstract) The study finds that women tend to use more words related to psychological and social processes, while men focus more on describing object features and impersonal topics. Although these differences are mostly context-dependent, the overall pattern indicates that gender differences are more pronounced in tasks with fewer limitations on language use. The AHRC-funded (AH/N002415/1) *Encyclopedia of Shakespeare's Language project* (2016-2019) "brings a new method of language research-the corpus approach-into the heart of Shakespearean studies. It affords fresh insights into Shakespeare's use of language at multiple levels-words, phrases, semantic themes, character profiles and more. In particular,

it reveals what Shakespeare's language meant to the Elizabethans through the analysis of millions of words written by his contemporaries." (*Encyclopedia of Shakespeare's Language*, Homepage) The two questions are pursued: "what does X mean" and "how is X used?"

This section examines Agatha Christie's use of linguistic elements in relation to her murderers and detectives. These linguistic elements are words identified in the previous section as carrying negative or positive sentiment scores. Two main research questions guide the analysis: (1) Which words with negative sentiment scores does Agatha Christie most frequently associate with *evil*, and which words with positive sentiment scores does she associate with *good*? (2) Does her use of these words differ between female and male murderers?

For clarity and comparability, adverbs which share the same stem with adjectives, are standardized into the adjective form. For instance, *carefully* and *careful* are standardized into one form: *careful*. Additionally, all verbs and nouns are presented in their lemmatized forms.

4.5.2 Result of Linguistic Representation

Modifiers with positive and negative sentiment scores are analyzed separately. For each sentiment category, a CSV file is generated with four columns: *words*, *role*, *gender*, and *frequency*. The purpose is to filter specific words, such as positive words associated only with murderers, negative words related exclusively to female characters or female murderers, and so on. Once the data is organized, the CSV file is imported into Gephi as a node table to create a visualization.

Figure 36 is a linguistic representation of positive modifiers in Gephi, with nodes colored according to gender. Orange nodes are modifiers used both for female and male characters, while red nodes indicate positive modifiers used exclusively for male characters and green nodes represent positive modifiers used exclusively for female characters. For all characters, the top three most frequently used positive words are: *thoughtful* (with a frequency of 13), *gently* (11), *smile* (10), followed by *agree* (7), *sympathetically* (6), *approve* (5) and *apologetically* (5), among others. Of all the 109 positive modifiers, 71.6%

are used for male characters and 28.4% are used for female characters. One reason for this can be attributed to the fact that five out of the eight works selected for analysis feature Poirot as the detective, which results in a higher overall number of modifiers associated with male characters. 41.28% of all positive modifiers are exclusively used for detectives, 33.94 % are used exclusively for murderers, 24.77% are used both for detectives and murderers, suggesting for positive sentiment, detectives are more strongly associated with positive language.

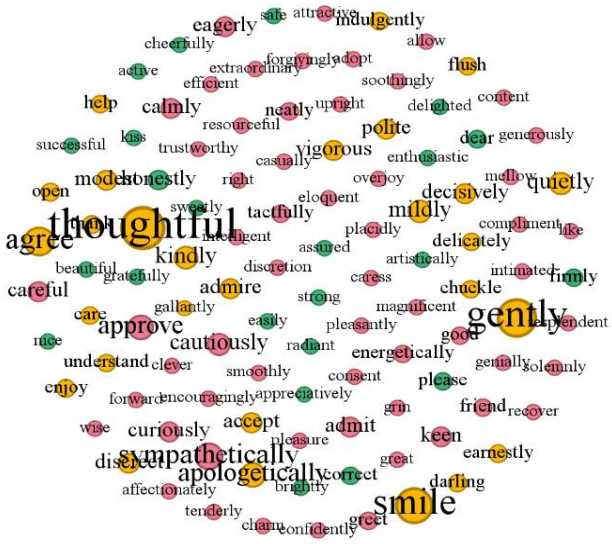

Figure 36: Linguistic Representation of Positive Modifiers in Gephi, Layout: Fruchterman Reingold

This study also focuses on the difference in language use between modifiers used for male and female murderers and Figure 37 represents the positive modifiers exclusively used for murderers.

There are 37 positive modifiers that are exclusively used for murderers. Among them, 14 words (37.84%) are exclusively used for female murderers and 18 words (48.65%) only for male murderers. The most frequently used modifiers for all murderers are *discreet* and *vigorous*, the most frequently and exclusively for female murderers used positive modifiers are *dear* and *please* and that for male murderers are *good* and *friend.*

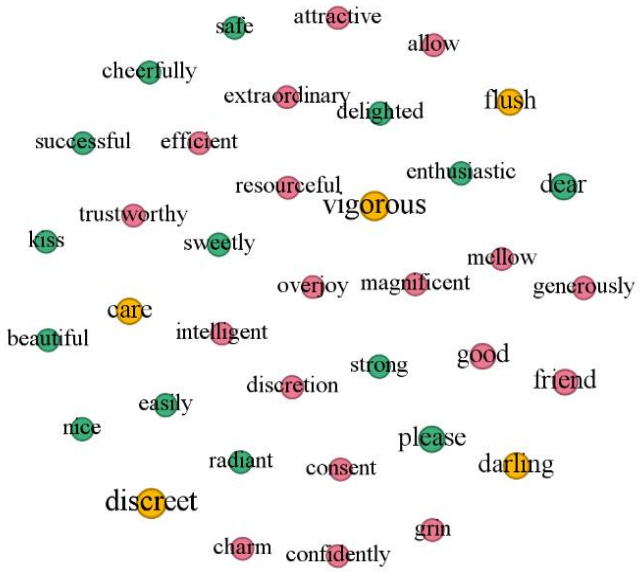

Figure 37: Linguistic Representation of Positive Modifiers Exclusively for Murderers in Gephi, Layout: Fruchterman Reingold

Positive modifiers exclusively used for female murderers and those used exclusively for male murderers are summarized in Table 4.

Male	**Frequency**	**Female**	**Frequency**	**Both**	**Frequency**
good	2	dear	2	discreet	3
friend	2	please	2	vigorous	3
attractive	1	safe	1	flush	2
allow	1	cheerfully	1		
efficient	1	kiss	1		
resourceful	1	beautiful	1		
trustworthy	1	nice	1		
overjoy	1	easily	1		
intelligent	1	sweetly	1		
discretion	1	delighted	1		
charm	1	radiant	1		
consent	1	enthusiastic	1		
confidently	1	strong	1		
grin	1	successful	1		

magnificent	1				
mellow	1				
generously	1				
extraordinary	1				

Table 4: Positive Modifiers Exclusively Used for Murderers

The following visualizations represent the negative modifiers. Figure 38 illustrates the negative modifiers for all characters, with nodes color-coded by gender: blue represents modifiers used for both genders, red represents those used exclusively for male characters, and green represents those used exclusively for female characters. For all characters, the top three most frequently used negative words are *cry* (with a frequency of 8), *interrupt* (7), *gravely* (6) and *coldly* (6), followed by *hesitate* (5), *frown* (5), *murder* (5), *sigh* (4), among others. Of all the 103 negative modifiers, 59.26% are used for male characters and 40.74% are used for female characters. One reason for this can still be attributed to the fact that five out of the eight works selected for analysis feature Poirot as the detective. 30.1% of all negative modifiers are exclusively used for detectives, 52.43 % are used only for murderers, 17.48% are used both for detectives and murderers. Within the domain of negative descriptors, murderers are clearly the primary targets.

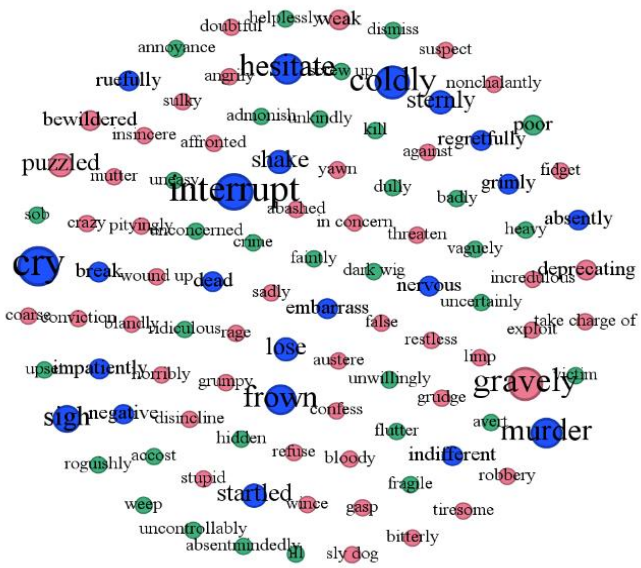

Figure 38: Linguistic Representation of Negative Modifiers in Gephi, Layout: Fruchterman Reingold

Figure 39 represents the negative modifiers exclusively used for murderers. There are 54 negative modifiers that are exclusively used for murderers. Among them, 23 words (42.59%) are exclusively used for female murderers and 27 words (50%) for male murderers. The most frequently used negative modifiers for all murderers are *murder* and *startled*, the most frequently and exclusively for female murderers used negative modifier is *poor* and that for male murderers is *weak.*

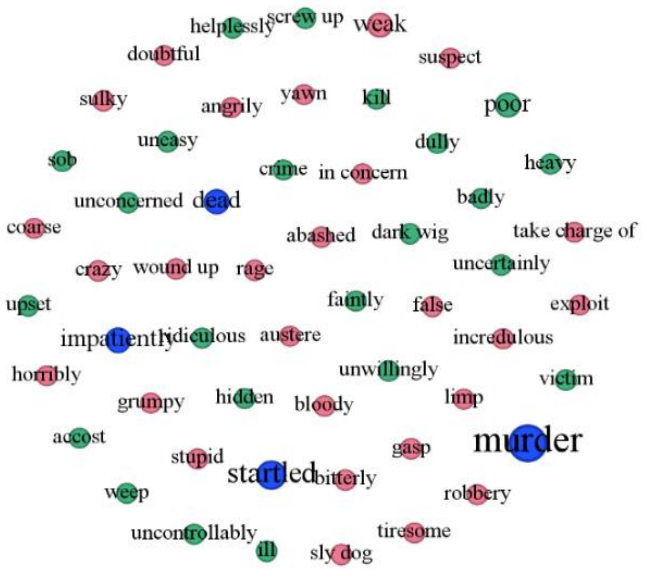

Figure 39: Linguistic Representation of Negative Modifiers Exclusively for Murderers in Gephi, Layout: Fruchterman Reingold

Negative modifiers exclusively used for female murderers and those used exclusively for male murderers are summarized in Table 5.

Male	**Frequency**	**Female**	**Frequency**	**Both**	**Frequency**
weak	2	accost	1	murder	5
yawn	1	badly	1	startled	3
wound up	1	crime	1	dead	2
tiresome	1	dark wig	1	impatiently	2
take charge of	1	dully	1		

suspect	1	faintly	1		
sulky	1	heavy	1		
stupid	1	helplessly	1		
sly dog	1	hidden	1		
robbery	1	ill	1		
rage	1	kill	1		
limp	1	ridiculous	1		
incredulous	1	screw up	1		
in concern	1	sob	1		
horribly	1	uncertainly	1		
grumpy	1	unconcerned	1		
gasp	1	uncontrollably	1		
false	1	uneasy	1		
exploit	1	unwillingly	1		
doubtful	1	upset	1		
crazy	1	victim	1		
coarse	1	weep	1		
bloody	1	poor	1		
bitterly	1				
austere	1				
angrily	1				
abashed	1				

Table 5: Negative Modifiers Exclusively Used for Murderers

Upon a close examination of the modifiers in Table 4 and Table 5, several conclusions can be drawn: regarding positive modifiers, male murderers are described more in terms of strength and intellect, with words like *resourceful*, *trustworthy*, *intelligent*, and *efficient*, while female murderers are more associated with emotional warmth, beauty, and charm, such as *safe*, *beautiful*, *kiss*, *radiant*, and *sweetly*. However, some modifiers typically associated with men in the author's era, like *strong* and *successful*, are also used to describe female murderers. Additionally, shared modifiers like *discreet* and *vigorous* suggest that murderers of both genders are seen as effective, energetic, or calculating. With respect to negative modifiers, male murderers are more described by a range of negative or problematic traits, such as emotional instability like *rage*, *grumpy*, *weak*, poor decision-

making like *stupid*, *sly dog*, or immoral behavior like *robbery*, *exploit*. This suggests a framing of male murderers in terms of instability, violence, or moral corruption**,** while the terms used for female murderers often convey a sense of passivity, emotional distress, or vulnerability, such as *sob*, *weep*, *helplessly*, *faintly*, *ill*, *poor*, and *upset*. However, modifiers exclusively used for female murderers like *dark wig*, *hidden*, and *victim* suggest that these female murderers adopt a facade of weakness to divert attention away from themselves during the investigative process. This result aligns with Dolski's (2024) argument that Agatha Christie's male murderers conform to hegemonic masculinity and societal expectations of men, while her female criminals combine traditionally masculine traits with feminine qualities, challenging gender binaries and societal norms, reflecting the concept of the "New Woman."

4.6 A Comprehensive Overview of the Murderers

In this section, the results of previous quantitative analyses are combined with additional information about the murderers in the selected novels, to create a comprehensive overview of them. Such information includes the murderer's age (old,>=60/middle-aged,40-60/young, <=40), occupations, motives, methods of killing, number of victims, and the murderer's fate (see Table 6).

	Gender	Age	Occupation/Identity	Motive	Methods of Killing	Number of Victims	Fate
Sheppard	male	middle-aged	Doctor	fear of exposure	stabbing	1	suicide
F. Clarke	male	young	a wealthy man's younger brother	greed	strangling, blunt force, stabbing	4	arrested
Wargrave	male	old	retired judge	twisted sense of justice	poison, blunt force, shooting and indirect	10	suicide

Knighton	male	young	secretary	greed	blunt force	1	arrested
P. Redfern	male	young	con artist	greed	blunt force	1	arrested
Jane Wilkinson	female	young	actress, Lord Edgare' wife	greed	stabbing, poison	3	arrested
Miss Blacklog	female	old	owner of a house	greed, fear of exposure	shooting, poison,strangling	3	arrested
Marina Gregg	female	young	actress	revenge, fear of exposure	poison, shooting	3	suicide
Ada Mason	female	young	house maid	greed	providing falsified testimony	1	arrested
C. Redferm	female	young	teacher of Latin	greed	providing falsified testimony	1	arrested

Table 6: Additional Information about the Murderers [9]

James Sheppard

James Sheppard is a middle-aged doctor in *The Murder of Roger Ackroyd*, serving as the private Doctor to the victim, Roger Ackroyd. Sheppard has been blackmailing Mrs. Ferrars, knowing that she has killed her husband. This immoral act becomes known to Roger Ackroyd, and out of fear of exposure, Sheppard murders him by stabbing him with a dagger. When his crime is eventually uncovered, Sheppard confesses and then takes his own life to save his sister from shame.

As the narrator, Sheppard is intensely active and communicative, shaping the reader's perspective while strategically concealing critical information. His social embeddedness is central to his deception. His respected standing as a modest village doctor, his longstanding friendship with Ackroyd, and his apparent cooperation with Poirot all ensure that he remains above suspicion. He even manages to redirect suspicion toward Ralph Paton, amplifying the shock of Poirot's final revelation. These narrative dynamics are strongly reflected in the SNA results: Sheppard exhibits elevated Degree, Closeness, Betweenness,

[9] The occupation/identity of the murderers only refer to their outward roles, e.g., Knighton's real identity is a jewel thief, but only his outward identity, the secretary to Rufus Van Aldin is relevant.

and Eigenvector centralities and high Quoted Speeches values, along with high speech/appearance rate. In contrast, his relatively low Mention value highlights how he maintains constant presence and influence while attracting minimal explicit attention, which is a key narrative strategy that helps him conceal his criminal role.

Sheppard is also the only murderer among the ten who displays a sentiment score of zero. This neutrality arises partly from his role as the narrator, which encourages a measured and controlled tone, but it also points to his underlying psychological profile. His actions are framed by self-preservation, rational calculation, and emotional restraint, producing an impression of clinical detachment. His linguistic behavior reinforces this effect: his narration frequently relies on understatement, omission, and neutral evaluative language, which together construct a façade of objectivity and reliability while strategically obscuring his involvement in the crime.

Franklin Clarke

Franklin Clarke is a young man, who has a very rich brother Sir Carmichael Clarke and Franklin Clarke would like to inherit his money. However, there is Miss Grey, for whom his brother has feelings. A possibility of their marriage would totally ruin Franklin's chances to inherit his brother's fortune thus he decides to kill his brother. Franklin kills three innocent people in an alphabetic order and makes the entire killings appear to be the acts of a madman. He uses different ways of methods in killing his victims: straggling a young lady, killing the first victim and his brother by use of blunt force, and stabbing the fourth victim with a knife. His sentiment score of -4.45 can be taken as a supporting point of his ruthless and cruel characteristic.

Franklin's method of committing murders in alphabetical order leads people to believe that the serial killings were carried out by a deranged lunatic. This works to Franklin's advantage, as he exploits Cust's mental instability, making Cust the most frequently mentioned and most suspicious person in the novel. Although Franklin appears later in the novel, his Degree Centrality, Closeness Centrality and Quoted Speeches are above average in SNA and his speech/appearance rate is at a high level, suggesting that he is active, communicative, confident, and even proud of his killing spree.

Franklin Clarke is arrested by the police in the end. Although he initially intends to commit suicide, Poirot intervenes and prevents him from doing so.

Justice Wargrave

Justice Wargrave is an old and retired judge and is obsessed by justice to an ill degree. He kills the 9 victims because they are not punished for their guilty. In the novel, there is a text describing Justice Wargrave and it comes from the victim Lombard: "…he's an old man and he's been presiding over courts of law for years. That is to say, he's played God Almighty for a good many months every year. That must go to a man's head eventually. He gets to see himself as all powerful, as holding the power of life and death–and it's possible that his brain might snap and he might want to go one step farther and be Executioner and Judge Extraordinary." (*And Then There were None*, 181)

Wargrave's murders are carried out in a methodical sequence, in accordance with the nursery rhyme "Ten Little Soldiers," employing various methods of killing, such as poisoning, blunt force, and shooting. He also manipulates Dr. Armstrong to stage his own death, ensuring that he would not be suspected as the murderer when the subsequent killings take place. Wargrave demonstrates confidence in his meticulously devised plan, utilizing his former position as a judge to incite suspicion among the victims, causing them to turn on each other. This ensures that, after his staged death, he does not need to personally kill the remaining victims: they will eliminate one another.

Wargrave's Degree Centrality, Quoted Speeches, Closeness Centrality, and Eigenvector Centrality values are all above average, indicating that he is well-connected to other characters and plays a highly active role in the novel. Despite this, his low Mention value reflects the fact that, as a former judge, no one initially suspects him. Wargrave appears frequently in the narrative but speaks relatively few on average during each appearance. In some scenes, he does not speak at all, which aligns with his judicial background: judges typically speak sparingly in the court, using their words with purpose and precision. When Wargrave does speak, he speaks extensively, *thoughtfully*, *calmly*, and *carefully*. His words consistently exert considerable influence over others, shaping their perceptions and actions, allowing him to *take charge of* the ongoing situation.

Wargrave takes his own life after all the victims are killed, driven by his obsession with justice: he believes that his own guilt in the murders should also be punished.

Major Knighton

Major Knighton is a young man working as the secretary to the millionaire Rufus Van Aldin. At the same time, he has another identity Le Marquis, the ruthless jewel thief. Major Knighton kills Rufus Van Aldin's daughter Ruth Kettering to steal her precious jewel by blunt force. With the help of his accomplice, Ada Mason, who has provided false statements to the police, and by leveraging his disguise as the *intelligent*, *trustworthy* secretary to Rufus Van Aldin, as well as his high sentiment score of 13.79, Major Knighton has successfully shifted suspicion onto Derek Kettering, Ruth's husband, and The Comet, Ruth's private lover. In SNA, all of Knighton's centrality values are above average; however, they remain well below the highest values. Combined with his low speech/appearance rate, Knighton is portrayed as a socially normal individual performing an ordinary job. Beneath this feigned normality and even quiet demeanor, however, lies his inconspicuous and dangerous nature. Ultimately, Knighton is arrested by the police.

Patrick Redfern

Patrick Redfern is a con artist as he likes *exploiting* women and is greedy of money, which leads to his strangle of wealthy Arlena Marschall, his private lover. With a positive sentiment score of 3.92, Patrick Redfern is portrayed as an *attractive*, *magnificent*, *good*, married young man. He is relatively active and communicative in the novel, maintaining close relationships with other characters. Patrick's high Mention score in SNA, while not the highest, indicates a tendency to seek attention from others. This portrayal suggests that Patrick is highly concerned with outward appearance and social impression, yet lacks the composure and rational restraint necessary for meticulous planning. His superficial charm and impulsive nature, reflected in negative modifiers such as *angrily*, *bitterly*, and *sulky*, makes him mentioned relatively often by others and renders him dependent on his wife, Christine, whose calmness and strategic thinking compensate for his deficiencies. It is precisely Christine's rationality and careful manipulation, such as creating alibis and

providing false testimony, that enable the murder plan to succeed and initially divert suspicion from the couple.

Jane Wilkinson

Jane Wilkinson is a young, *beautiful,* and *successful* actress, who thinks only of herself. She is cruel and selfish, which can also be reflected in her sentiment score -11.77. She wants to divorce Lord Edgware to marry the Duke of Merton. However, a normal divorce does not satisfy her, because the Duke of Merton will not wed a divorcée and she also wants to inherit a big fortune from his husband. So, she brutally kills Lord Edgware by stabbing the back of his head. Since Carlotta Adams is utilized by Jane Wilkinson to create an alibi and Donald Ross recognizes her conspiracy in a party, Jane Wilkinson kills them by poisoning and stabbing respectively. She is arrested and sent to prison, waiting for her execution in the end.

Jane Wilkinson is active and communicative throughout the novel which is reflected through her high centrality values in SNA and high speech/appearance rate. She is also the most mentioned figure. Combining all the features together, it can be revealed that Jane Wilkinson is representative of the concept of the "New Woman": she is not subdued to man, since she does not accept Lord Edgware's admission of divorce; she is intelligent, since she utilizes people to carry out her murder plan; she represents her power with male traits, since she brutally killed her husband and Donald Ross by stabbing; she is the most frequently mentioned character, since she does not live and behave within societal expectation. In the end, in her confession, she even hopes that her execution would be different from the usual: she wants to be hanged in public.

Miss Blacklog

Miss Blacklog is an elderly lady in her sixties who owns a house where other people also live. She paid Rudi Scherz to organize a murder party with the intent to kill him, because Rudi discovered that Miss Blacklog has falsified her identity. The real name of Miss Blacklog is Charlotte Blacklog, but she assumes the identity of her deceased sister, Letitia Blacklog to inherit a large sum of money from Letitia's former employer. Fearing exposure,

Miss Blacklog devises this murder plot and stages it to appear as though she herself is the intended victim.

In the novel, Miss Blacklog is described as a *cheerful*, *gentle*, and *vigorous* lady, always *willing to help,* with a positive sentiment score of 4.55. Her high centrality values (except for Betweenness Centrality) and her moderately high speech/appearance level suggests that she is relatively active and communicative, maintaining close relationships with others. Miss Blacklog is portrayed as a harmless, poor person, and due to her fortune obtained through impersonating her sister Letitia, no one suspects her to be the true target of the crime. However, Miss Blacklog is far from a simple old woman. She is highly intelligent but driven by greed and an intense sense of insecurity. Cold and calculating, she kills her good friend, Dora Bunner, with poison and later strangles Miss Murgatroyd, all to eliminate suspicion and protect her secret. Miss Marple stopped her fourth murder of the house maid, Mitzi. In the end, Miss Blacklog is arrested.

Marina Gregg

Marina Gregg is a beautiful, famous, and young actress. Like Miss Blacklog, who is also portrayed positively (with a sentiment score of 4.08), Marina Gregg has staged a murder to create the illusion that she is the intended victim. Marina kills Mrs. Badcock out of revenge because Mrs. Badcock unknowingly transmitted German measles to Marina, leading to a miscarriage.

At the cocktail party, Mrs. Badcock accidentally spilled her cocktail, and Marina gave Mrs. Badcock her own drink in exchange. This is the drink that is poisoned. As a result, Mrs. Badcock drank the poisoned cocktail and died. Since the drink originally belongs to Marina, she has successfully staged the murder to make it appear as though she is the intended victim.

After the first murder, Marina claims to have been shocked and ill from the incident, which has caused her to stay at home. This is why her appearances in the novel are concentrated in the first and last few chapters, resulting in her low centrality values. However, Jane Wilkinson belongs to the highly talkative murderers: her each appearance is accompanied by a large number of quoted speeches. These speeches might be her way of conducting a

deviant investigation into the case, trying to manipulate others' perceptions or to confuse the investigation.

Marina is not only a character with a high average speech frequency but also the most mentioned character in the story. This is tied to her being seen as the real target of the murder and her status as a public female figure. Like Miss Blacklog, while pretending to be shocked and frightened, Marina continues to kill two more victims who might have discovered her secret and are blackmailing her: one is shot, and the other is poisoned. In the end, Marina chooses to commit suicide with poison. (There is an alternative theory: she might have asked her husband to help her end her life, and it is he who has administered the poison, helping her escape the consequences of her crimes.)

Ada Mason

As the young housemaid of the victim in *The Mystery of the Blue Train*, Ada Mason is a quiet and seemingly peripheral character, reflected through her slightly above-average Degree Centrality and low Closeness Centrality in SNA. Due to her *uncertainty* and *upset* during the interrogation, her low frequency of appearance in the story, a small number of quoted speeches in the story and no co-occurrence with the main perpetrator, Major Knighton, no one suspects her of being an accomplice. She is mostly seen as a frightened little maid (with a below zero sentiment score). However, as Agatha Christie herself mentioned in her autobiography, all the clues are in the text. For example, during the interrogation, Ada Mason often *discreetly* coughs, which is her signal for creating a false testimony.

When it is revealed that Ada Mason is an accomplice, and that her real identity is a male impersonator and actress, all the doubts are cleared. For example, she disguised herself as the victim, Ruth Kettering, to make Derek Kettering believe that her wife, facing away from him, lying in bed, was still alive when he entered Ruth's compartment. Mason also disguised herself as a man after handling Ruth's body and left the train. It is noteworthy that Mason has a moderately high speech/appearance rate, suggesting that she has intensively engaged in the interrogation, i.e., providing falsified testimony. Ada Mason's connection to Major Knighton is not mentioned directly in the story, but it is later explained in Poirot's resolution of the case.

Ada Mason does not directly commit the murder, but without her false testimony and the disruption caused by impersonating other people, the murder, which is aimed at stealing jewels, would not have been completed. The fate of Ada Mason is not explicitly mentioned in the novel. However, given that Poirot and the police have discovered her real identity, it can be assumed that her fate involves arrest.

Christine Redfern

In the *Evil under the Sun*, Christine Redfern is not only the wife of the main perpetrator Patrick Redfern but also his accomplice. Christine is a teacher of Latin and young. She pretends to be hurt by the romantic relationship between her husband and the victim Arlena Marschall and is portrayed as an emotionally weak person (with a below zero sentiment score), like *ridiculous*, *helplessly*, *sob*, *faintly*, *screw up*, *ruefully*. However, under her disguise, she has helped his husband by providing falsified testimony.

As a couple, the Redferns share the same circle of social relationships, thus, they share almost the same metric values in SNA. Christine's speech/appearance is slightly higher than that of Patrick Redfern. This is her way of helping her husband. As mentioned before, Patrick's superficial charm and impulsive nature make him prone to exposure, it is Christine's rationality and careful manipulation that enable the murder plan to succeed and initially divert suspicion from the couple. The Redferns are arrested for their crime.

The murderers in these selected novels exhibit very different features. They come from various occupations, belong to different age groups, and have distinct social relationships. The motives for their crimes also vary, including revenge, fear of exposure, greed, and a distorted sense of justice. The methods of murders are not consistent either: some opt for poison, others use blunt force, while some employ firearms or knives. These findings align with Verbogen's (2008) argument that Agatha Christie balances formulaic elements with inventive variations with respective to her conception of murderers.

In the eight novels examined, the majority of the murderers are ultimately arrested and three choose to commit suicide. Notably, the characters who choose suicide, Sheppard, Justice Wargrave, and Marina Gregg, are all fully aware of their guilt prior to ending their lives. Sheppard hopes that by taking his own life, he could protect his sister from the public scandal. Wargrave believes that, having killed others, he too deserves to face punishment.

Similarly, Marina Gregg ends her life to stop her own descent into madness. This suggests that Agatha Christie's treatment of evil is unequivocal: evil must be terminated. As she has written in her autobiography "the detective story was the story of the chase: it was also very much a story with a moral; in fact, it was the old Everyman Morality Tale, the hunting down of Evil and the triumph of Good." (*Agatha Christie An Autobiography*, 437)

Agatha Christie's portrayals of female murderers reflect her moderate feminism, particularly through the concept of the "New Woman." These five female murderers all divert suspicion by exploiting society's ingrained stereotypes of women. Ada Mason pretends to be a weak housemaid, Christine Redfern feigns emotional damage from her husband's extramarital affair, both Miss Blacklog and Marina Gregg present themselves as the intended victims of murders, and Jane Wilkinson acts as though she is a dramatic actress who could not hide any secrets. However, these women use their underlying intelligence to plan their murders or to fabricate credible false testimonies for their male accomplices. Their methods of killing are not limited to poisoning: they also exhibit the same violence as their male counterparts, such as stabbing or strangling.

There is an interesting phenomenon: each of the three female solo murderers is the most frequently mentioned character in her respective narrative. They are the seemingly least likely individuals to commit the crime, yet they are the most frequently mentioned persons, suggesting that, while they are playing on their societal stereotypes, they do not redirect suspicion onto others. This can, to some extent, reflect their inner satisfaction and confidence in the murder scenarios they have orchestrated. They believe that they do not need to accuse any specific person to ensure their crimes remain undiscovered; they are capable of creating a perfect crime scene.

Finally, the female murderers in this thesis span a chronological range: Ada Mason in 1928, Jane Wilkinson in 1933, Christine Redfern in 1941, Miss Blacklog in 1950, and Marina Gregg in 1962. This progression suggests that, over time, Agatha Christie's feminist views did not undergo a significant change.

5. Conclusion and Outlook

In this thesis, methodologies from Digital Humanities are applied to characterize the murderers from eight of Agatha Christie's detective novels: social network analysis reflects the relationships the murderers have with other characters based on co-occurrence; bar plots of speech/appearance rate reflect their level of talkativeness; Word2Vec contextualization aims to find whether there are distinct boundaries between murderers and non-murderers; sentiment analysis uncovers whether all the murderers are as negative represented in the novels; linguistic representation of the murderers provides a further insight into the linguistic description of them. These digital methods are employed as a supporting and supplementary tool to close reading, reflecting Agatha Chritie's conception of murderers and her moderate feminism.

Through quantitative analysis, it becomes clear that Agatha Christie's murderers are complex and multifaceted, reflecting the inherent unreadability of her texts. Some murderers are depicted as seemingly positive, while others are overtly negative. Some are highly active and loquacious, while others remain reserved and silent. Some are frequently mentioned by other characters, drawing attention and suspicion, while others divert suspicion onto others through minimal mention. There are diverse motives and methods for the killings, further complicating their characterizations. However, in the end, the fate of all her murderers is either arrest or suicide, underscoring Agatha Christie's clear stance on evil: it must be terminated.

By analyzing female murderers in Agatha Christie's novels, it becomes evident that they are active and communicative, often maintaining close relationships with other characters. While they are frequently depicted as emotionally vulnerable, poor, harmless, and lacking in intelligence, these portrayals serve as disguises to carry out their murder schemes and divert police investigations. These women employ the same violent methods as their male counterparts, such as stabbing and strangling, and are depicted with similar traits of discretion and vigor. Female accomplices also play a crucial role, offering strategic false testimonies to support the male perpetrators. Together, these elements suggest that these female murderers are not only intelligent but also assertive and independent, challenging traditional gender expectations. These nuanced portrayals align with Agatha Christie's moderate feminism: the "New Woman."

This thesis represents a novel attempt to explore the characterizations of Agatha Christie's murderers through the lens of Digital Humanities. The analyses present results that validate its theoretical framework, such as the previously discussed unreadability of her texts and the concept of "New Woman." As a master thesis, which is constrained by both time and scope, it is not feasible to expand the corpus as much as one might desire. Therefore, only eight selected works from Agatha Christie's extensive bibliography are analyzed. In the future, the use of Digital Humanities methods could be extended to a larger corpus of texts, which allows for more robust results and clearer patterns. Moreover, future research could build upon the findings by conducting more focused and nuanced analyses using similar methods. For instance, in SNA, it could be valuable not only to expand the textual corpus but also to refine the representation of relationships between murderers and other characters. These connections could be examined in greater details, such as through their simultaneous presence in specific locations or dialogue scenes. Moreover, directed relationships could be explored, capturing social hierarchies (e.g., superior-subordinate, father-daughter, husband-wife, doctor-patient) and emotional dynamics (e.g., love, hatred, neutrality, sympathy, respect). In addition, future linguistic analyses could investigate gender-based differences in representation. For example, whether female murderers employ distinctive morphological or syntactic patterns.

Overall, this thesis demonstrates that quantitative analysis can support and complement insights gained from close reading, showing how data-driven approaches can enrich traditional literary interpretation by bridging computational methods and interpretive scholarship.

6. Literature

6.1 Primary Literature

A Murder is Announced: https://detective.gumer.info/anto/christie_19_2.pdf

And Then There were None: https://detective.gumer.info/anto/christie_117_2.pdf

Evil under the Sun: https://detective.gumer.info/anto/christie_58_2.pdf

Lord Edgware Dies: https://detective.gumer.info/anto/christie_42_2.pdf

The A.B.C. Murders: https://detective.gumer.info/anto/christie_43_2.pdf

The Mirror Cracked Side from Side: https://detective.gumer.info/anto/christie_36_2.pdf

The Murder of Roger Ackroyd: https://www.gutenberg.org/cache/epub/69087/pg69087.txt

The Mystery of the Blue Train: https://www.gutenberg.org/cache/epub/72824/pg72824.txt

6.2 Secondary Literature

Agarwal, Apoorv, Auguto Corvalan, Jacob Jensen, and Owen Rambow. "Social Network Analysis of *Alice in Wonderland*." *Proceedings of the NAACL-HLT Workshop on Computational Linguistics for Literature*, Association for Computational Linguistics, Montréal, Canada. 2022, pp. 88-96.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, und Tomas Mikolov. "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics*, vol. 5, 2017, pp. 135-146. DOI: 10.1162/tacl_a_00051.

Burrows, John F. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method.* Clarendon Press, 1987.

Bourdieu, Pierre. "Ökonomisches kapital - kulturelles kapital - soziales kapital." Kreckel R.(ed.), *Soziale Ungleichheiten*, Kreckel R., Schwarz, Göttingen, 1983, pp. 183-198.

Cahyadi, Ni Made Dewi, and Dede Brahma Arianto. "Sentiment Analysis of Characters *Romeo and Juliet* by Shakespeare Using Naïve Bayes Algorithm." *Journal of Applied Linguistics (ALTICS)*, vol. 6, no. 1, 2024, pp. 41–51. https://doi.org/10.36423/altics.v6i1.1596

Christie, Agatha. *Agatha Christie An Autobiography*. 5th ed., HarperCollins Publishers, 2011.

Dolski, Dora. "The portrayal of male and female murderers in selected novels of Agatha Christie." University of Zadar, 2024. URN: nbn:hr:162:247069

Domjanović, Vedran, and Biljana Oklopčić. "Murder and Crime-Solving Strategies in Agatha Christie's Works." *Folia Linguistica et Literaria*, vol. 39, 2022. DOI:10.31902/fll.39.2022.1

Drakman, Amnelie, and Stefan Gelfgren. "How to combine close and distant reading within the history of science and ideas: Two examples from ongoing research." Lärdomshistoriska Samfundet, 2022, pp. 85-108. DOI: 10.48202/24128

Eder, Maciej. "Visualization in Stylometry: Cluster Analysis Using Networks." *Digital Scholarship in the Humanities*, vol. 32, no. 1, 2015, pp. 50-64.

Firth, John Rupert. “A Synopsis of Linguistic Theory 1930-55.” *Studies in Linguistic Analysis*, The Philological Society, 1957, pp. 1–32.

Fischer, Frank and Daniil Skorinkin. “Social Network Analysis in Russian Literary Studies.” Gritsenko, Daria, Mariëlle Wijermars and Mikhail Kopotev (eds), *The Palgrave Handbook of Digital Russia Studies,* Palgrave Macmillan, Cham, 2021, pp. 517-536. https://doi.org/10.1007/978-3-030-42855-6_29

Gius, Evelyn, Svenja Guhr, and Inna Uglanova. “‘d-Prose 1870–1920’: A Collection of German Prose Texts from 1870 to 1920.” *Journal of Open Humanities Data*, vol. 7, no. 11, 2021. DOI:10.5334/johd.30.

Görmar, Maximillian, Jörn Münkner, and Hartmut Beyer. “Digitale Exploration und hermeneutische Bewertung: Profilierung einer frühneuzeitlichen Gelehrtenfigur mittels Netzwerkanalyse am Beispiel von Leonhard Christoph Sturm (1669-1719).” *Zeitschrift für digitale Geisteswissenschaften*, 2022. DOI: 10.17175/2022_008_v2

Jansen, Dorothea. *Einführung in die Netzwerkanalyse*, Springer Fachmedien, Wiesbaden, 2003.

Jockers, Matthew**.** *Macroanalysis: Digital Methods and Literary History.* University of Illinois Press, 2013.

Justyna, Patyk. “A stylometric analysis of detective fiction and other works of Agatha Christie, Arthur Conan Doyle and Gilbert Keith Chesterton.” Jagiellonian University, 2022. https://ruj.uj.edu.pl/xmlui/handle/item/302980

Hark, Ina Rae. “Impossible Murderers: Agatha Christie and the Community of Readers.” Delamater, Jerome H. and Ruth Prigozy (eds), *Theory and Practice of Classic Detective Fiction,* Bloomsbury Publishing USA,1997, pp. 111-118.

Makinen, Merja. *Agatha Christie: Investigating Femininity*. Palgrave Macmillan, 2006.

Min, Sungkyu, and Juyong Park. “Modeling Narrative Structure and Dynamics with Networks, Sentiment Analysis, and Topic Modeling.” *PLOS ONE*, vol. 14, no. 12, 2019. https://doi.org/10.1371/journal.pone.0226025.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. “Efficient Estimation of Word Representations in Vector Space.” arXiv:1301.3781v3 [cs.CL], 2013. https://doi.org/10.48550/ARXIV.1301.3781.

Moretti, Franco. “Conjectures on World Literature.” *New Left Review*, no. 1, Jan.-Feb. 2000, pp. 54-68. DOI:10.64590/hxj.

Moretti, Franco. *Distant Reading*. Konstanz University Press, Konstanz, 2011.

Najar, Esmaeil, and Fatemeh Vaziri. “Hercule Poirot and Criminal Psychology: Crime and Detection in Selected Novels of Agatha Christie.” *Journal of Language Horizons*, Alzahra University, 2021. DOI: 10.22051/lghor.2020.32304.1344

Nalisnick, Eric T. and Henry S. Baird. “Extracting Sentiment Networks from Shakespeare’s Plays.” *12th International Conference on Document Analysis and Recognition*, 2013, pp. 758-762.

Newman, Mathew L., Carla J. Groom, Lori D. Handelman, and James W. Pennebaker. “Gender Differences in Language Use: An Analysis of 14,000 Text Samples.” *Discourse Processes*, 45(3), 2008, pp. 211-238. https://doi.org/10.1080/01638530802073712

Pichler, Axel, and Nils Reiter. “From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities.” *Journal of Cultural Analytics*, vol. 7, no. 4, Dec.2022. https://doi.org/10.22148/001c.57195

Poole, Robert. *Corpus-Assisted Ecolinguistics*. London, Bloomsbury Publishing, 2022.

Stiller, James, Daniel Nettle, and Robbin I.M Dunbar. "The Small World of Shakespeare's Plays." *Human Nature*, vol. 14, no. 4, 2003, pp.397-408.

Tamanna. "Mapping Melancholy: Sentiment Analysis of Emotional Trends in Victorian Literature." *Shodhaamrit*, vol. 2, no. 1, 2025, pp. 20-28. https://doi.org/10.71037/Shodhaamrit.v2i1.02

Theodar, Caleb. "Fractured Whodunits, Reconstructed Selves: Computational Narratives of Hercule Poirot Across Page and Screen." SRM Institute of Science and Technology, 2025. SSRN: http://dx.doi.org/10.2139/ssrn.5379036 .

Trilcke, Peer. "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft." Ajouri, Phillip, Katja Mellmann, and Christoph Rauen (eds), *Empirie in der Literaturwissenschaft*, Münster, 2013, pp. 201-247.

Verbogen, Aagje. "Agatha Christie and Her Murderers: A Case Study of the Miss Marple Novels." University of Gent, 2008. https://libstore.ugent.be/fulltxt/RUG01/001/414/024/RUG01-001414024_2010_0001_AC.pdf

Wasserman, Stanley and Katherine Faust. *Social Network Analysis: Methods and Application*. Cambridge University Press, 1994. https://doi.org/10.1017/CBO9780511815478

Winckles, Andrew O., and Angela Rehbein, editors. *Women's Literary Networks and Romanticism: "A Tribe of Authoresses."* Liverpool University Press, 2017. https://doi.org/10.2307/j.ctt1ps32xn

Winterkvist, Frida. "There Is Nothing More Deceptive than an Obvious Fact: A Feminist Study of the Detective Work by Miss Marple and Sherlock Holmes." University of Gävle, 2020. URN: urn:nbn:se:hig:diva-32473

Young, Lori and Stuart Soroka. "Affective News: The Automated Coding of Sentiment in Political Texts." *Political Communication*, vol. 29, no. 2, 2012, pp. 205-231. DOI:10.1080/10584609.2012.671234.

Zeiger, Carolyn Allen. *The Miss Marple Model of Psychological Assessment*, Burns Park Publishers, 2013.

6.3 Tools

Anthony, Laurence. AntConc (Version 3.2.1). Tokyo, Japan, Waseda University, 2023. http://www.antlab.sci.waseda.ac.jp/

Bastian, M., S. Heymann, and M. Jacomy. Gephi: an open-source software for exploring and manipulating networks. *Third international AAAI conference on weblogs and social media*,2009. https://gephi.org/

DeepL Translator. DeepL SE, 2025. https://www.deepl.com/translator

Fischer, Frank and Carsten Milling. Easy Linavis (Simple Network Visualisation for Literary Texts). HDH2017: "Sociedades, políticas, saberes". Málaga. Libro de resúmenes, 18–20 October 2017. pp. 173-176. https://ezlinavis.dracor.org/

Gius, Evelyn, Jan Christoph Meister, Malte Meister, Marco Petris, Dominik Gerstorfer, and Mari Akazawa. CATMA 7 (Version 7.1), 2024. https://catma.de/

OpenAI. ChatGPT (GPT-5). 2025. https://chat.openai.com

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2023. https://www.r-project.org/

Schumacher, Mareike. “forTEXT-Lerneinheit Word2Vec.” *GitHub*, 2023. https://github.com/forTEXT/forTEXT.net/blob/main/word2vec/LerneinheitWord2Vec.ipynb

Young, Lori and Stuart Soroka. *Lexicoder Sentiment Dictionary*. 2012. https://www.snsoroka.com/data-lexicoder/

6.4 Internet Access

Encyclopedia of Shakespeare's Language Project. Lancaster University. https://wp.lancs.ac.uk/shakespearelang/contact/ (Last access: Oct.25,2)

Everything Agatha. https://everythingagatha.com. (Last access: Oct.21, 2025)

Schumacher, Mareike. “Netzwerkanalyse.” *forTEXT: Literatur digital erforschen*, 2018, https://fortext.net/routinen/methoden/netzwerkanalyse. (Last access: Oct. 20, 2025)

Schumacher, Mareike. “CATMA.” *forTEXT: Literatur digital erforschen*, 2019, https://fortext.net/tools/tools/catma. (Last access: Oct. 20, 2025)

https://en.wikipedia.org/wiki/The_Murder_of_Roger_Ackroyd (Last access: Oct.20,2025)

https://en.wikipedia.org/wiki/And_Then_There_Were_None (Last access: Oct.20,2025)

https://en.wikipedia.org/wiki/The_A.B.C._Murders (Last access: Oct.20,2025)

https://en.wikipedia.org/wiki/A_Murder_Is_Announced (Last access: Oct.20,2025)

https://en.wikipedia.org/wiki/Lord_Edgware_Dies (Last access: Oct.20,2025)

https://en.wikipedia.org/wiki/The_Mirror_Crack%27d_from_Side_to_Side (Last access: Oct.20,2025)

https://en.wikipedia.org/wiki/The_Mystery_of_the_Blue_Train (Last access: Oct.20,2025)

https://en.wikipedia.org/wiki/Evil_Under_the_Sun (Last access: Oct.20,2025)

6.5 List of Figures and Tables

Figure 1: Workflow for Reflected Text Analysis, Pichler & Reiter (2022)

Figure 2: Visualization of Character Relationships in *The Murder of Roger Ackroyd* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 3: Visualization of Character Relationships in *The Murder of Roger Ackroyd* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

Figure 4: Visualization of Character Relationships in *The Murder of Roger Ackroyd* with Nodes Ranked by Mention, Layout: Yifan Hu

Figure 5: Visualization of Character Relationships in *The A.B.C. Murders* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 6: Visualization of Character Relationships in *The A.B.C. Murders* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

Figure 7: Visualization of Character Relationships in *The A.B.C. Murders* with Nodes Ranked by Mention, Layout: Yifan Hu

Figure 8: Visualization of Character Relationships in *And Then There were None* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 9: Visualization of Character Relationships in *And Then There were None* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

Figure 10: Visualization of Character Relationships in *And Then There were None* with Nodes Ranked by Mention, Layout: Yifan Hu

Figure 11: Visualization of Character Relationships in *Lord Edgware Dies* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 12: Visualization of Character Relationships in *Lord Edgware Dies* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

Figure 13: Visualization of Character Relationships in *Lord Edgware Dies* with Nodes Ranked by Mention, Layout: Yifan Hu

Figure 14: Visualization of Character Relationships in *A Murder is Announced* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 15: Visualization of Character Relationships in *A Murder is Announced* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

Figure 16: Visualization of Character Relationships in *A Murder is Announced* with Nodes Ranked by Mention, Layout: Yifan Hu

Figure 17: Visualization of Character Relationships in *The Mirror Cracked Side from Side* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 18: Visualization of Character Relationships in *The Mirror Cracked Side from Side* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

Figure 19: Visualization of Character Relationships in *The Mirror Cracked Side from Side* with Nodes Ranked by Mention, Layout: Yifan Hu

Figure 20: Visualization of Character Relationships in *The Mystery of the Blue Train* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 21: Visualization of Character Relationships in *The Mystery of the Blue Train* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

Figure 22: Visualization of Character Relationships in *The Mystery of the Blue Train* with Nodes Ranked by Mention, Layout: Yifan Hu

Figure 23: Visualization of Character Relationships in *Evil under the Sun* with Nodes Ranked by Degree, Layout: Yifan Hu

Figure 24: Visualization of Character Relationships in *Evil under the Sun* with Nodes Ranked by Quoted Speeches, Layout: Yifan Hu

Figure 25: Visualization of Character Relationships in *Evil under the Sun* with Nodes Ranked by Mention, Layout: Yifan Hu

Figure 26a: Bar Plot of Speech/Appearance Rate of Murderers

Figure 26b: Bar Plot of Speech/Appearance Rate of Murderers

Figure 27: Semantic Filed of Miss Blacklog with Top 20 Most Similar Words

Figure 28: t-SNE Representation of Semantic Fields of Murders and Non-Murderers individually

Figure 29: t-SNE Representation of Semantic Fields of Murders and Non-Murderers in Group

Figure 30: A Screenshot of Cluster for the Murderer Miss Blacklog in AntConc

Figure 31: A Screenshot of KWIC Result for the Word Sequence *Blacklog said* in AntConc

Figure 32: A Screenshot of *Blacklog*'s Modifiers in CSV Format

Figure 33: A Screenshot of Sentiment Analysis for the Modifiers of Miss Blacklog

Figure 34-35: Sentiment Network Visualizations for Miss Blacklog and Miss Marple in Gephi, Layout: Yifan Hu

Figure 36: Linguistic Representation of Positive Modifiers in Gephi, Layout: Fruchterman Reingold

Figure 37: Linguistic Representation of Positive Modifiers Exclusively for Murderers in Gephi, Layout: Fruchterman Reingold

Figure 38: Linguistic Representation of Negative Modifiers in Gephi, Layout: Fruchterman Reingold

Figure 39: Linguistic Representation of Negative Modifiers Exclusively for Murderers in Gephi, Layout: Fruchterman Reingold

Table 1: Synthesis of Network Analyses of Murderers' Measurement Scores in Selected Novels

Table 2: Murderers' Average Number of Speeches per Chapter of Appearance

Table 3: Sentiment Scores of Murderers and Detectives in Selected Novels

Table 4: Positive Modifiers Exclusively Used for Murderers

Table 5: Negative Modifiers Exclusively Used for Murderers

Table 6: Additional Information about the Murderers

6.6 Repository

https://github.com/VivienMengjiao/Master-Essay

Selbstständigkeitserklärung

Ich bestätige hiermit, dass ich von den Plagiatsregeln am Institut für Literaturwissenschaft Kenntnis genommen habe und erkläre, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel verwendet und alle den verwendeten Quellen und Hilfsmitteln wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ich erkläre hiermit ferner, dass ich die KI-/IT-gestützten Schreibwerkzeuge DeepL Translator /OpenAI ChatGPT (GPT-5) in allen Kapiteln zum Zwecke der Rechtschreibprüfung/Verbesserung der Formulierung/Übersetzung chinesischer Wörter ins Englisch. Darüber hinaus wurde OpenAI ChatGPT (GPT-5) zur Generation von Code für den statistischen t-test, die t-SNE Visualisierung (Figuren in Gruppen) eingesetzt. Weitere KI-basierten Hilfsmittel habe ich nicht verwendet.

Ort, Datum

Unterschrift