

Agatha Christie Across Time: A Textual Analysis

影片連結 : <https://youtu.be/ptYA6FDB0QQ>

第 13 組

B104020002 劉和媛

B104020006 莊子儀

B104020008 林冠儒

B104020037 董昱辰



00 Table of Contents

→ 01

Introduction

→ 02

Sentiment
Analysis

→ 03

Topic
Exploration

→ 04

Relation
Graph

01

Project

Introduction

01 Introduction: Motives

本專案將探討推理女王 -- *Agatha Christie* 的作品。由於 *Agatha Christie* 不僅是人類史上最暢銷的作家之一（作品銷量僅次於聖經與莎士比亞），且作品豐富、橫跨多個年代，同時一些作品如《東方快車謀殺案》、《尼羅河謀殺案》等更是被翻拍成影視作品。因此我們想探討 隨著年代進展(如二戰前後)、主角更換(白羅 or 瑪波小姐)等其作品會有什麼變化？



01 Introduction: Dataset

我們從Github、z-library蒐集1920-1979之Agatha Christie小說，並將EPUB檔轉換為TXT檔以進行分析。

年份(年)	數量(本)
1920-1929	11
1930-1939	24
1940-1949	14
1950-1959	14
1960-1969	11
1970-1979	9



02

Sentiment Analysis

02 Sentiment Analysis

本部分想探討 *Agatha Christie* 作品中的情緒變化模式，並按年份將作品劃分成多個時期，分析比較不同時期作品的敘事情感特徵。

分析工具：

- 主要使用 NLTK VADER 情緒分析方法
- 搭配機器學習的 TextBlob 工具進行輔助辨別
- 建立包含 21 個謀殺相關關鍵詞的詞庫
- 結果視覺化呈現

```
murder_keywords =  
['murder', 'killed', 'death',  
'dead', 'corpse', 'body',  
'victim', 'poison', 'shot',  
'stabbed', 'strangled',  
'blood', 'crime', 'suspect',  
'detective', 'investigation',  
, 'kill', 'die', 'fatal',  
'assassin', 'homicide' ]
```

1920-1929

資料集概況

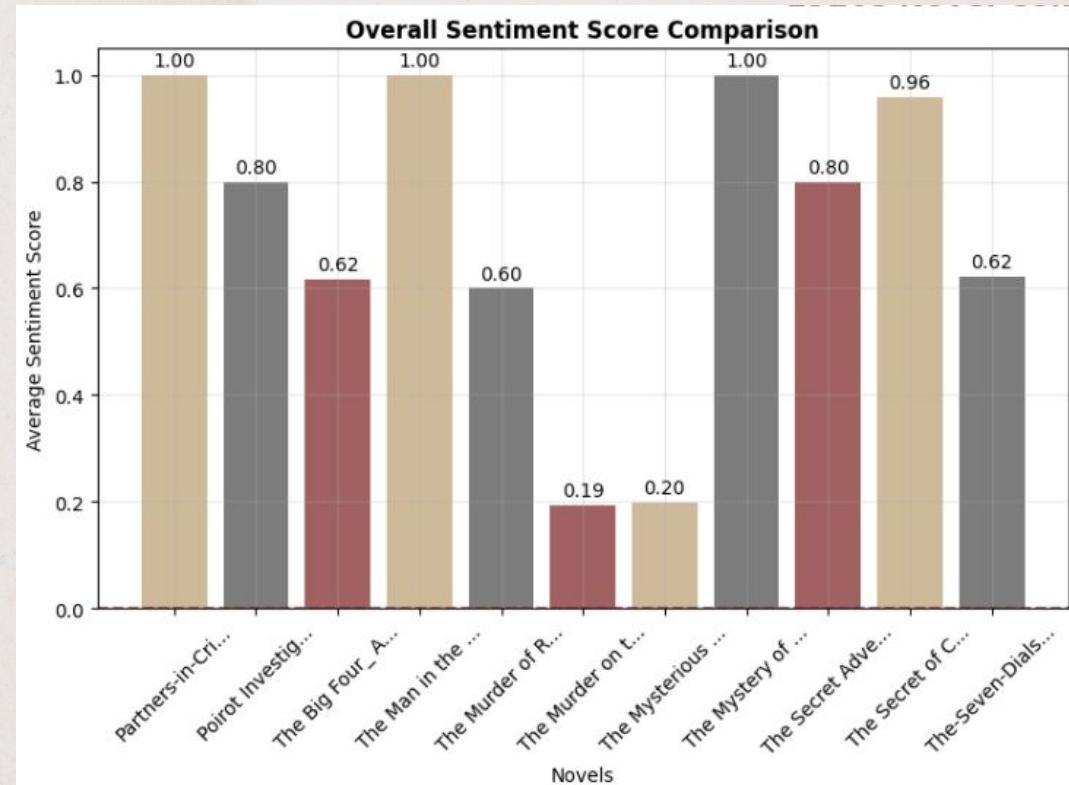
- 分析小說數量: 11 本
- 總字數: 757,361
- 平均字數: 68,851

分析結果

- 平均情緒分數: 0.708
- 平均情緒波動: 0.535

黃金時代的經典謎案

顯示在 1920 年代的寫作風格傾向營造 **相對樂觀** 或**輕鬆**的敘事氛圍，即使在處理謀殺主題時也能維持某種優雅感。



1930-1939

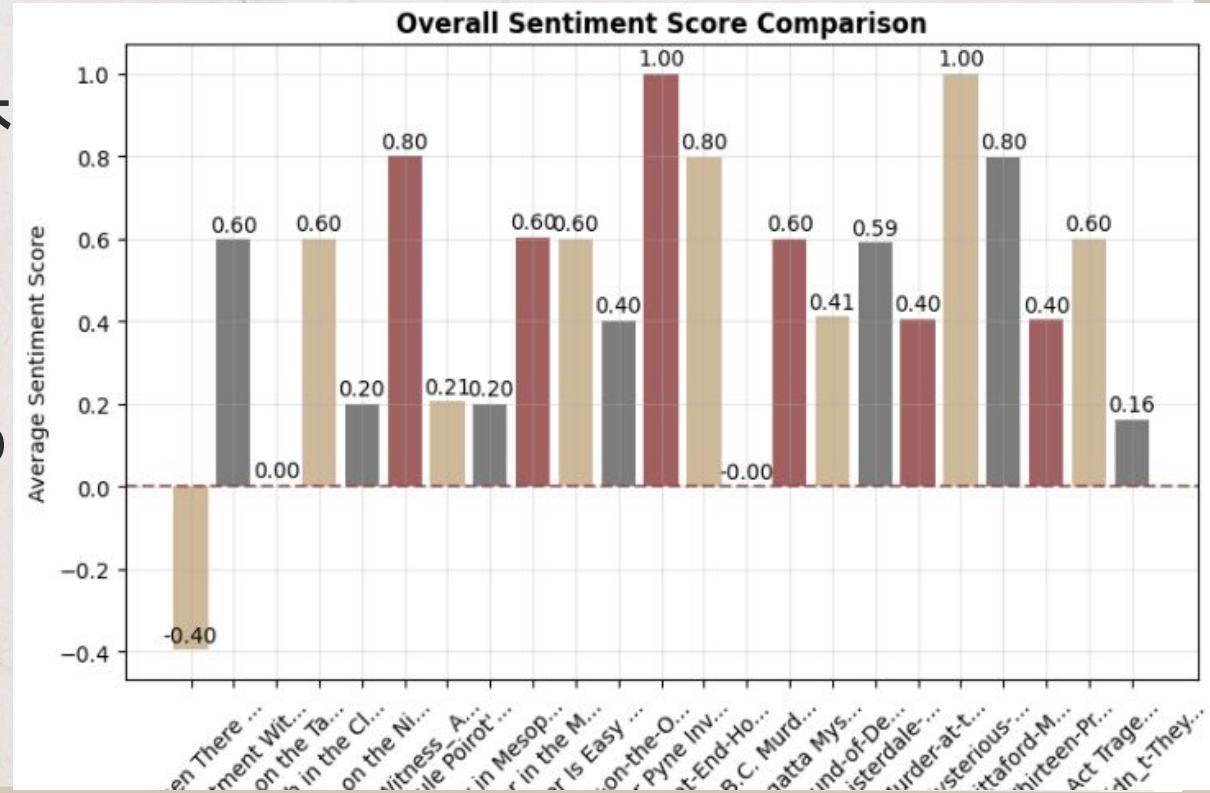
資料集概況

- 分析小說數量: 23 本
- 總字數: 1,342,060
- 平均字數: 58,350

分析結果

- 平均情緒分數 : 0.460
- 平均情緒波動 : 0.810

出現明顯的負面情緒作品(如:《 There Was an Old Woman》的-0.40),能反映作者開始嘗試不同創作風格或更深層心理主題的探索。



1940-1949

資料集概況

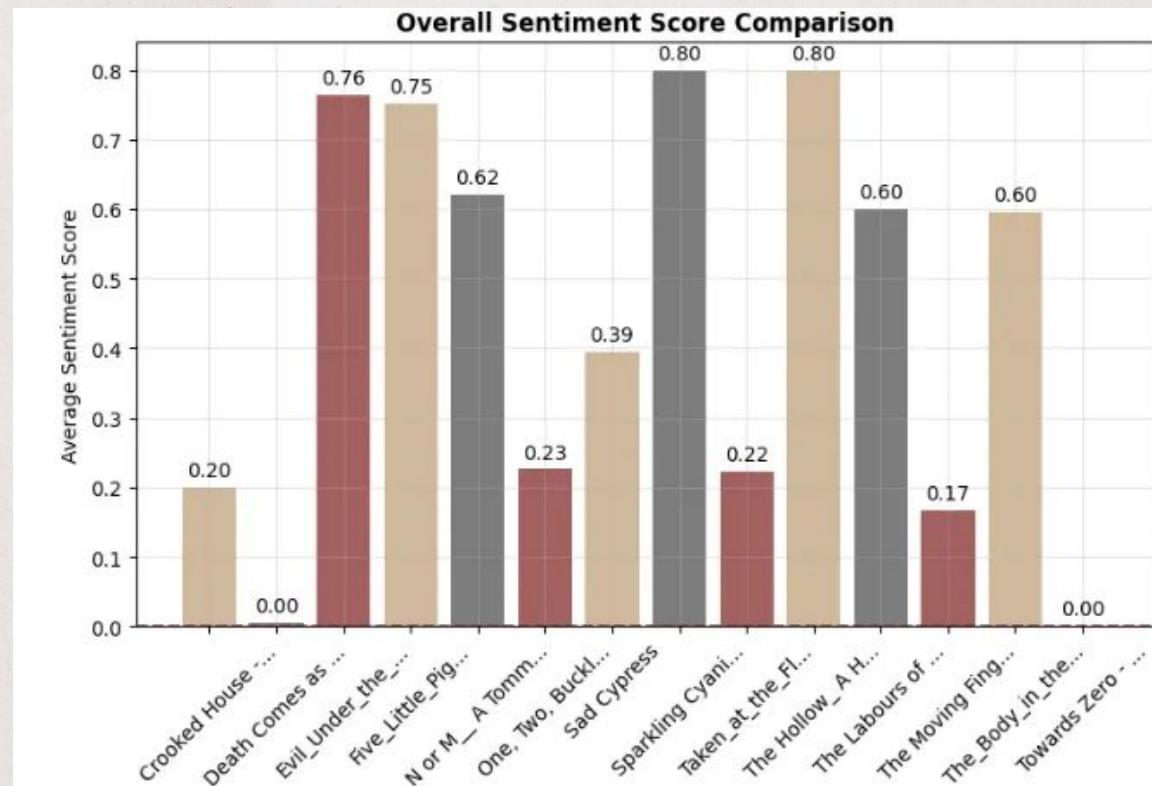
- 分析小說數量: 14 本
- 總字數: 849,727
- 平均字數: 60,695

分析結果

- 平均情緒分數: 0.439
- 平均情緒波動: 0.864

戰爭陰影與心理深度

出現更多中低情緒分數的作品(如:《N or M?》的0.23、《Sparkling Cyanide》的0.22),可能反映了對人性陰暗面和戰爭時期社會心理的更深入探討。



1950-1959

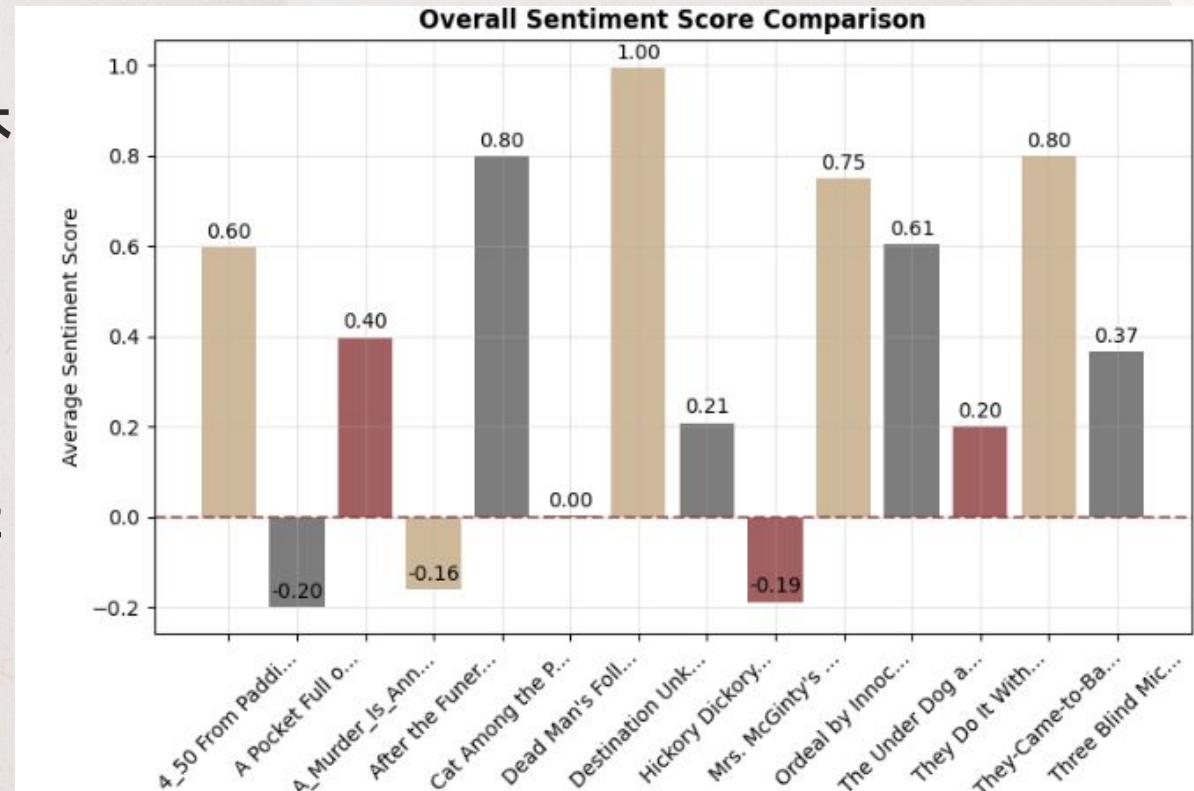
中期作品中負面情緒的增加可能也反映 當時社會環境的變化，或是作者對人生和社會更深層次的思考。

資料集概況

- 分析小說數量：14 本
- 總字數：888,844
- 平均字數：63,489

分析結果

- 平均情緒分數：0.370
- 平均情緒波動：0.832



1960-1969

在接近創作晚期時，趨向於 避免過度陰暗的敘事基調，可能反映了某種創作智慧的積累。

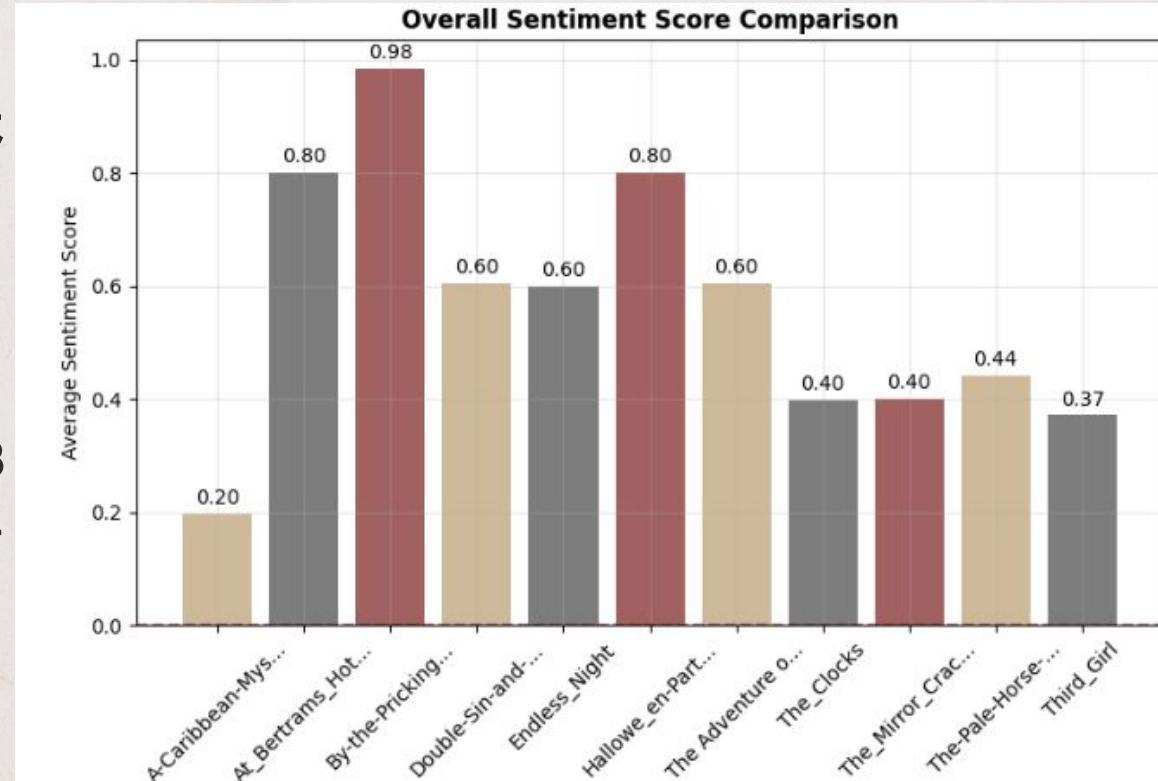
資料集概況

- 分析小說數量: 11 本
- 總字數: 726,145
- 平均字數: 66,013

分析結果

- 平均情緒分數 : 0.563
- 平均情緒波動 : 0.784

衰退期與實驗性嘗試



1970-1979

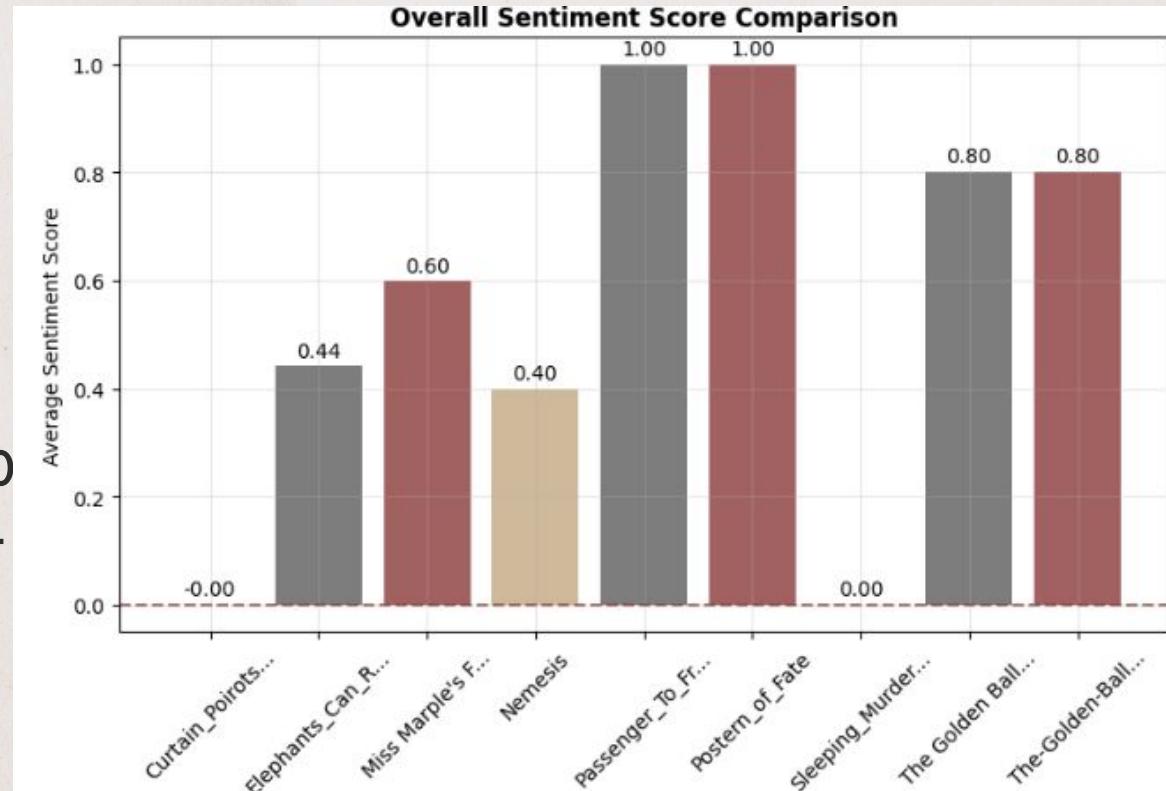
資料集概況

- 分析小說數量: 9 本
- 總字數: 603,610
- 平均字數: 67,068

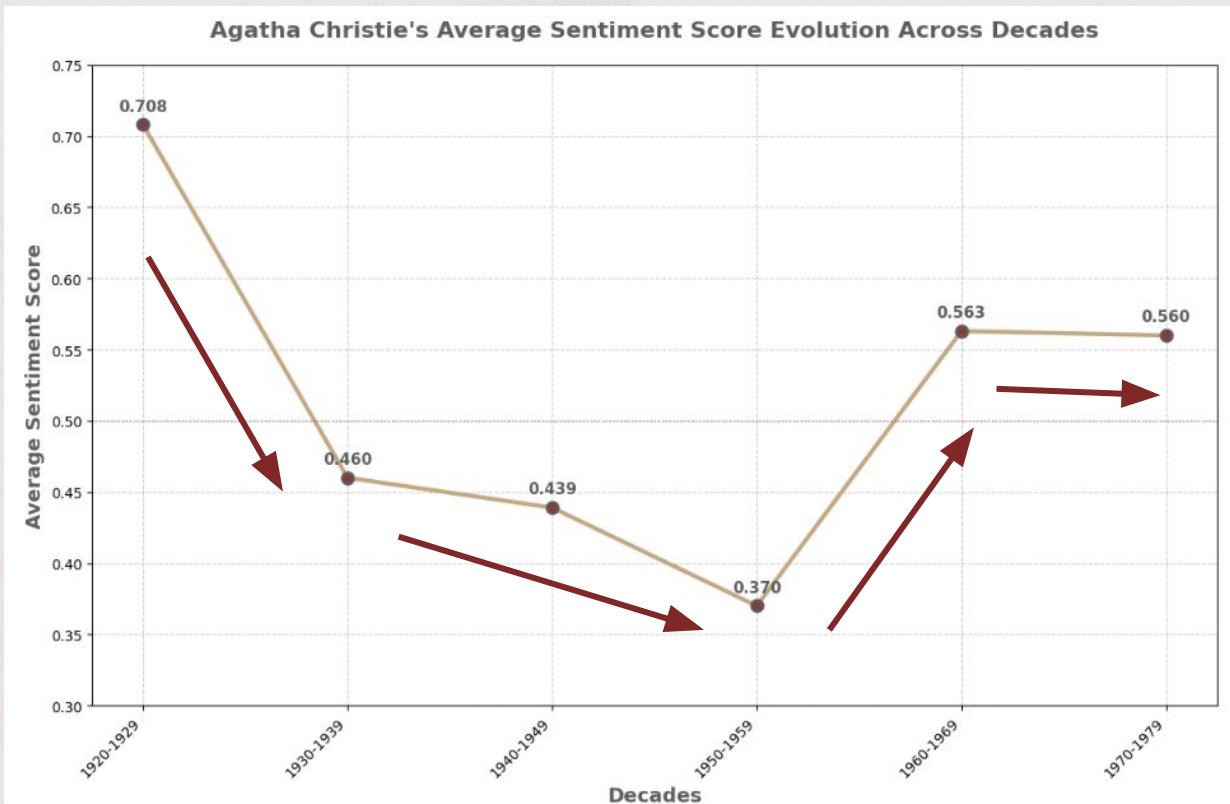
分析結果

- 平均情緒分數: 0.560
- 平均情緒波動: 0.674

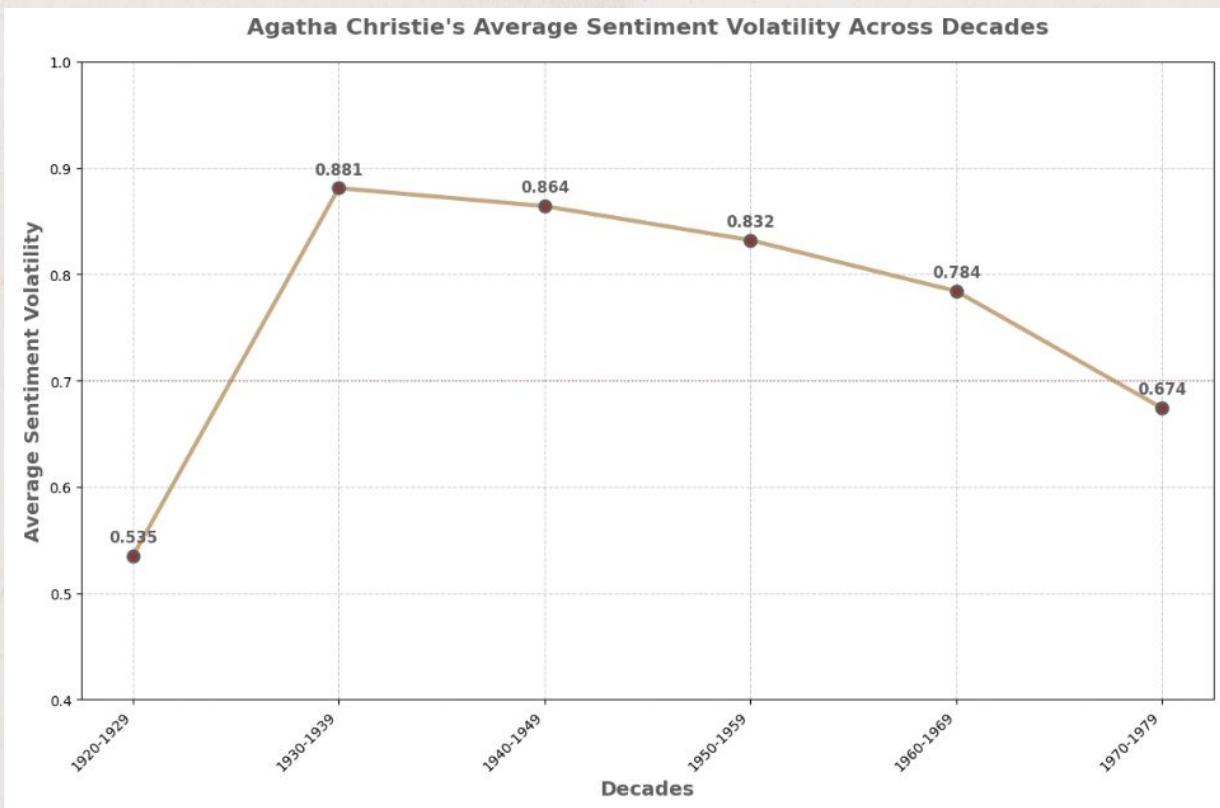
最後階段的情緒分布反映了作者 對人生、創作和告別的深層思考，為整個創作生涯畫下了意味深長的句號。



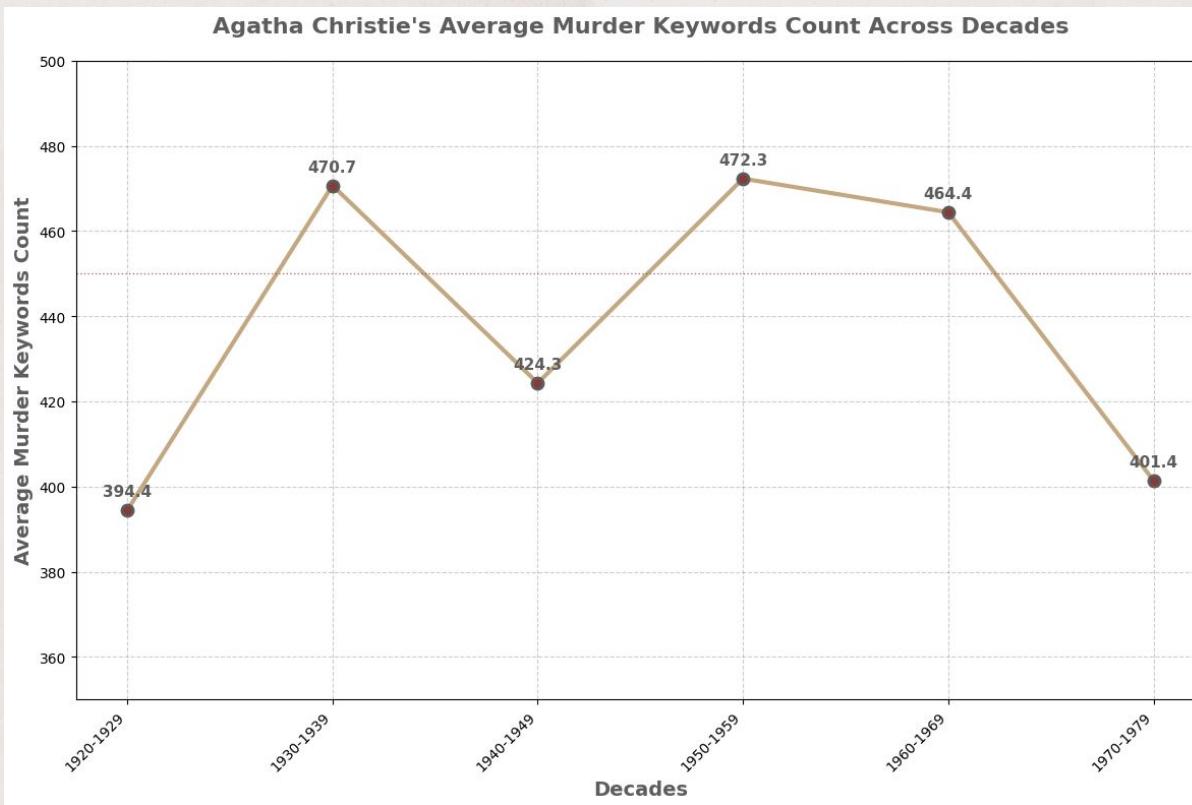
02 Sentiment Analysis: Average Score Evolution Across Decade



02 Sentiment Analysis: Average Volatility Evolution Across Decade



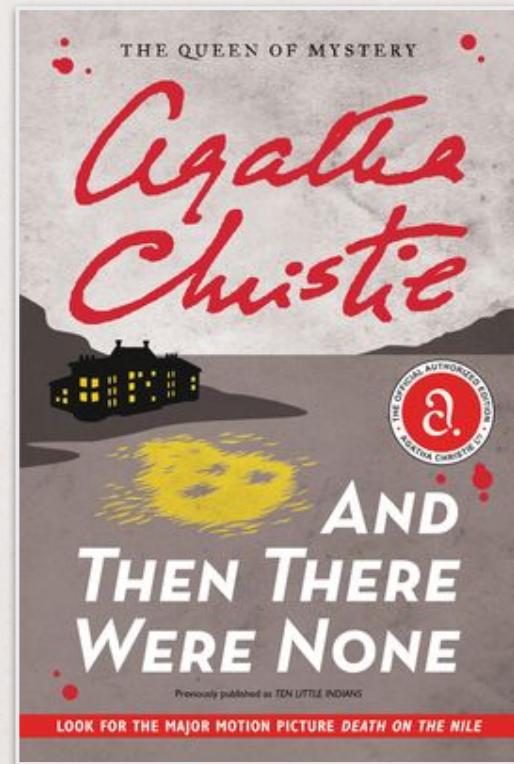
02 Sentiment Analysis: Average Murder Keywords Count Across Decades



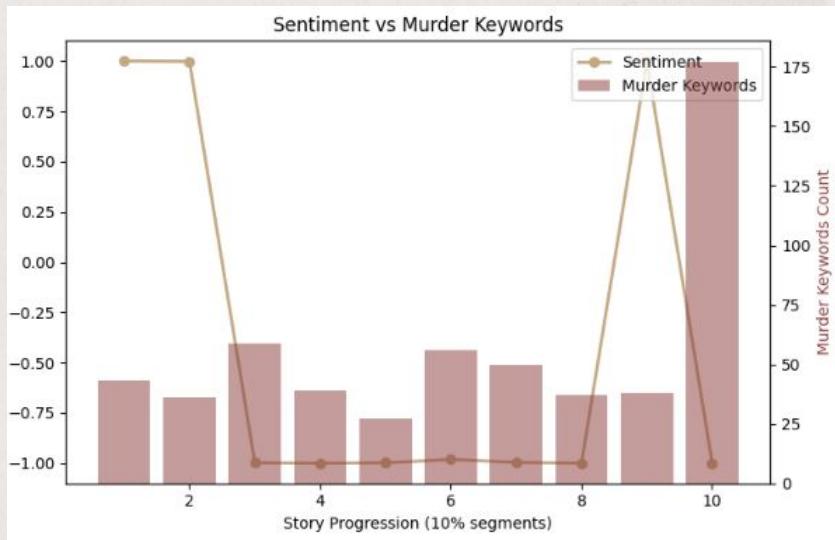
02 Sentiment Analysis: 單一文本深入分析

And Then There Were None - 1939

Agatha Christie 於1939年出版的經典推理小說，故事講述十個陌生人受邀至孤島，神秘主人未現身。島上留聲機指控每人犯罪，隨後按童謠順序逐一死亡。兇手藏身其中，營造極度恐懼氛圍，為史上最偉大推理小說之一。

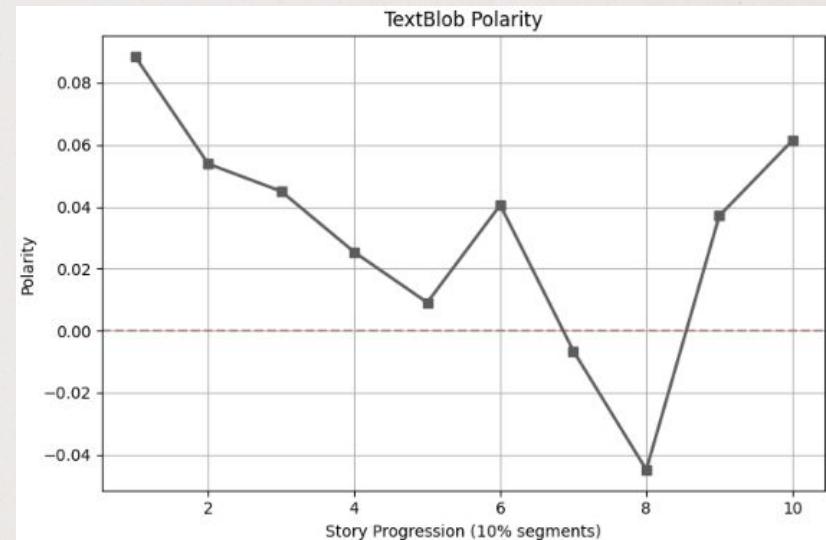


02 Sentiment Analysis: 單一文本深入分析



Sentiment Pattern

- 前8段：持續負面情緒，營造陰暗氛圍
- 第9段：情緒劇烈飆升，形成強烈對比
- 第10段：敘述真相，回歸平靜情緒



Insights into literary techniques

- 懸疑營造：前期穩定負面情緒建立緊張感
- 高潮設計：第9段可能為 **真相揭露** 或 **關鍵轉折點**
- 情緒操控：先抑後揚的經典推理小說結構

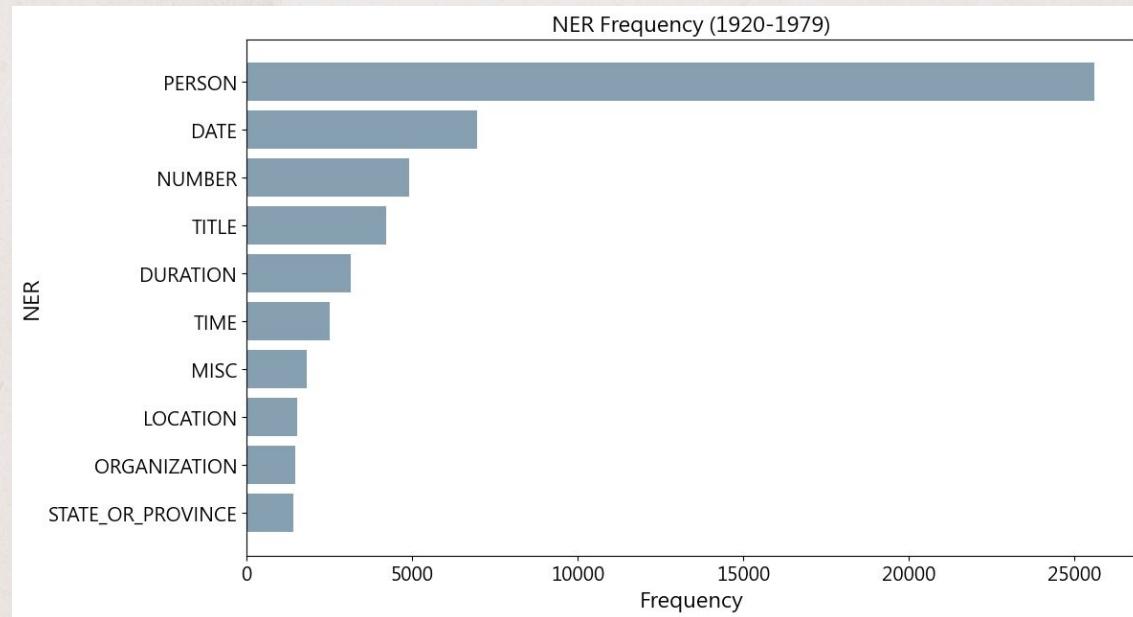
03

Topic
Exploration

03 Topic Exploration: 1920-1979 CoreNLP NER

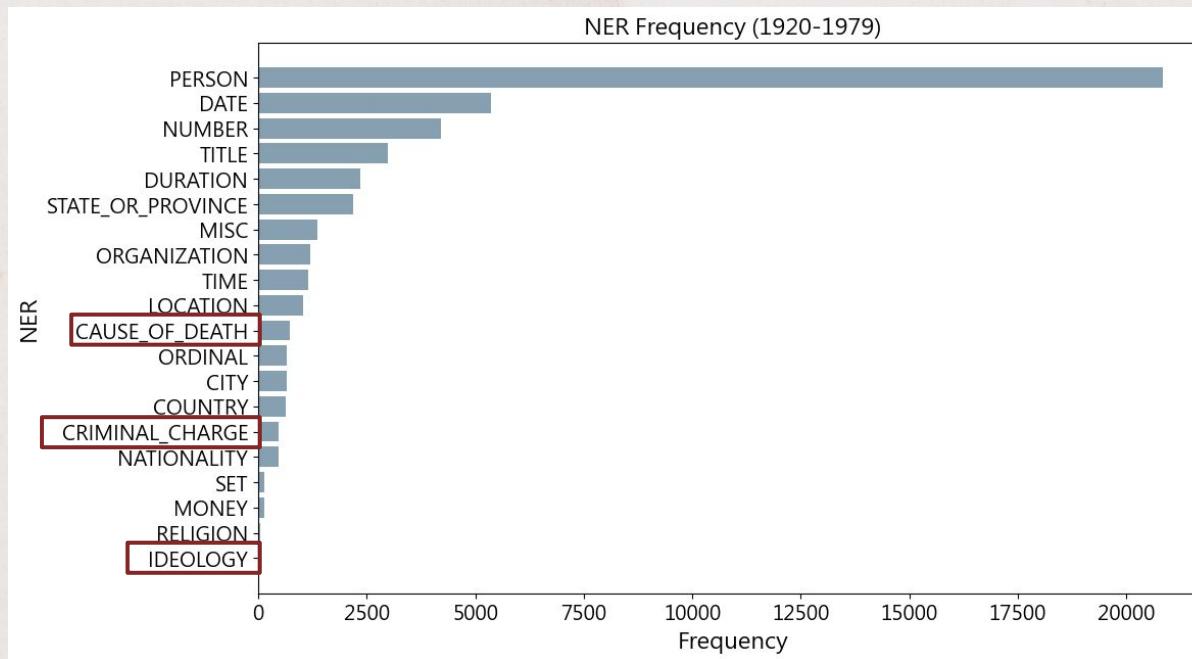
小說核心元素

- **PERSON**: 偵探、嫌疑人、受害者、證人
- **DATE**: 案件發生的時間
- **TITLE**: 人物的職稱、頭銜
- **DURATION**: 案發時長
- **TIME**: 案發時間
- **LOCATION**: 案發地點



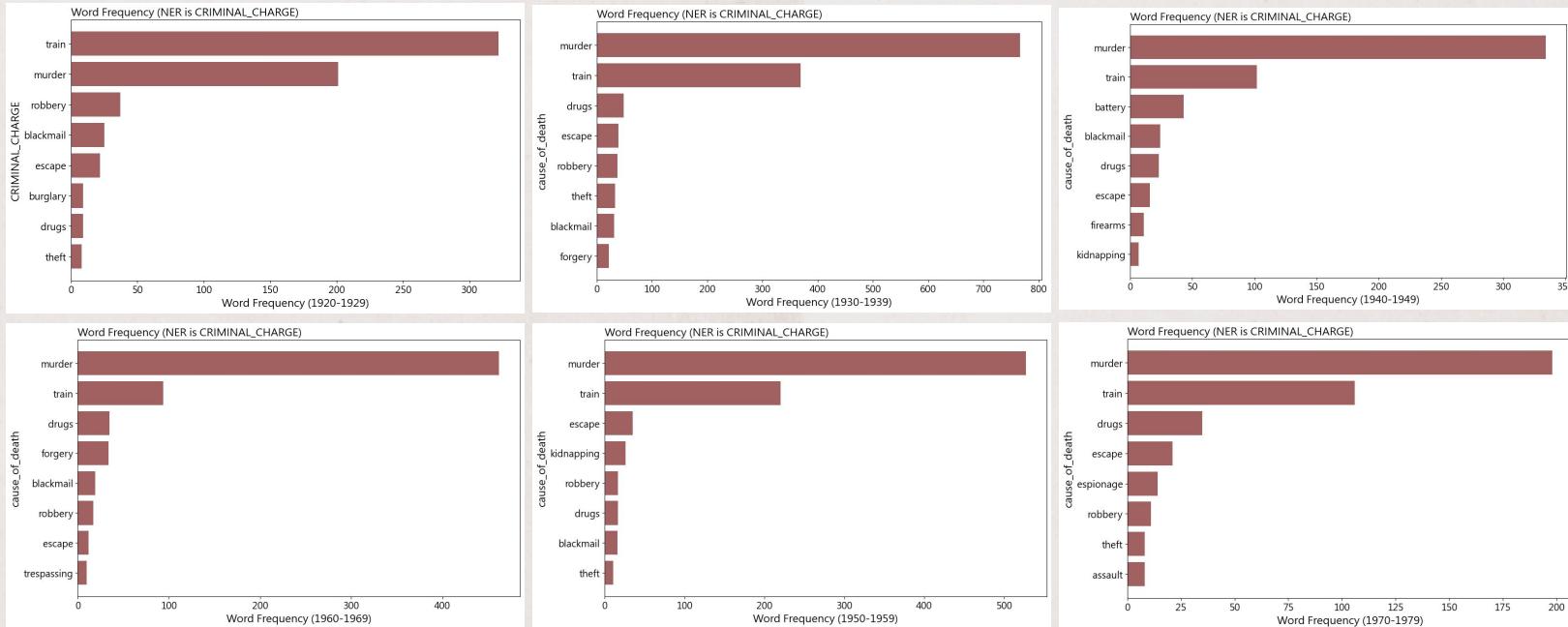
出現在文本中前十多的實體，主要圍繞在故事中的 人物、時間、地點。主要強調角色之間的互動、案件發生的時間線，以及場景背景的描寫等特徵。不同年代的實體排序變化不大，反映出小說的基本組成結構大致一致。

03 Topic Exploration: 1920-1979 CoreNLP NER



選擇 Cause Of Death、Criminal Charge、Ideology，進一步觀察。

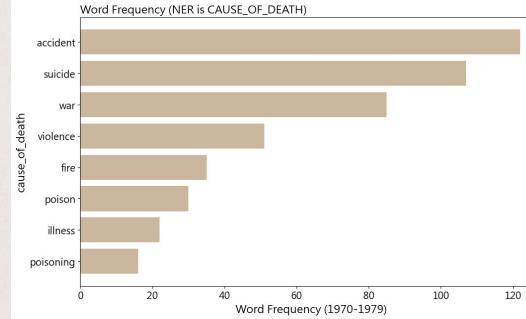
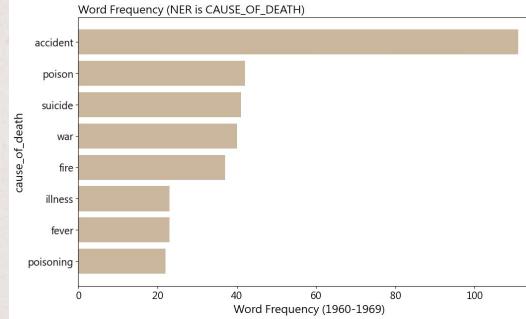
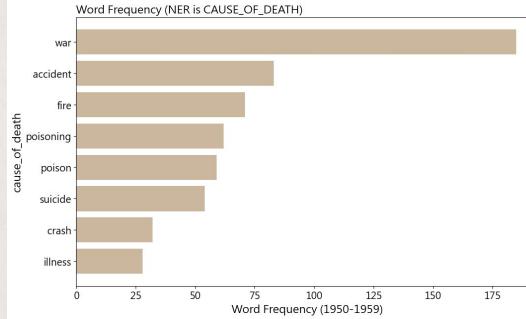
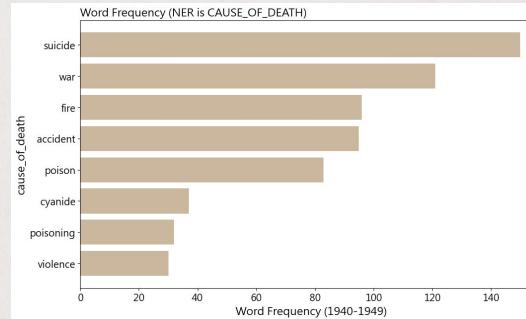
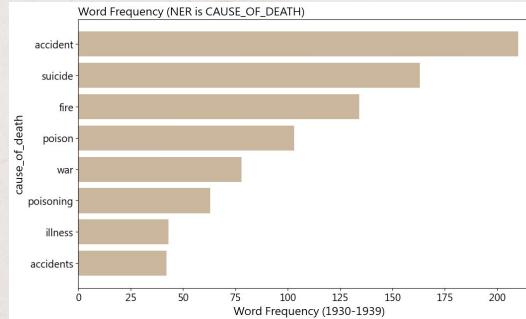
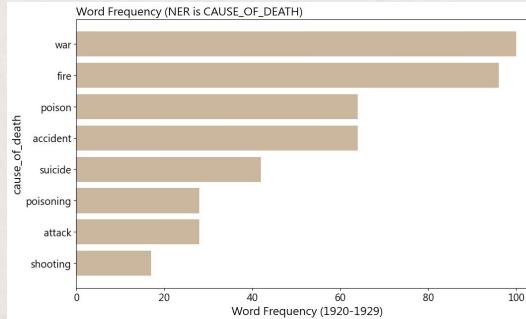
03 Topic Exploration: NER_Criminal Charge



在不同年代中皆以 **murder** 出現次數為最多。

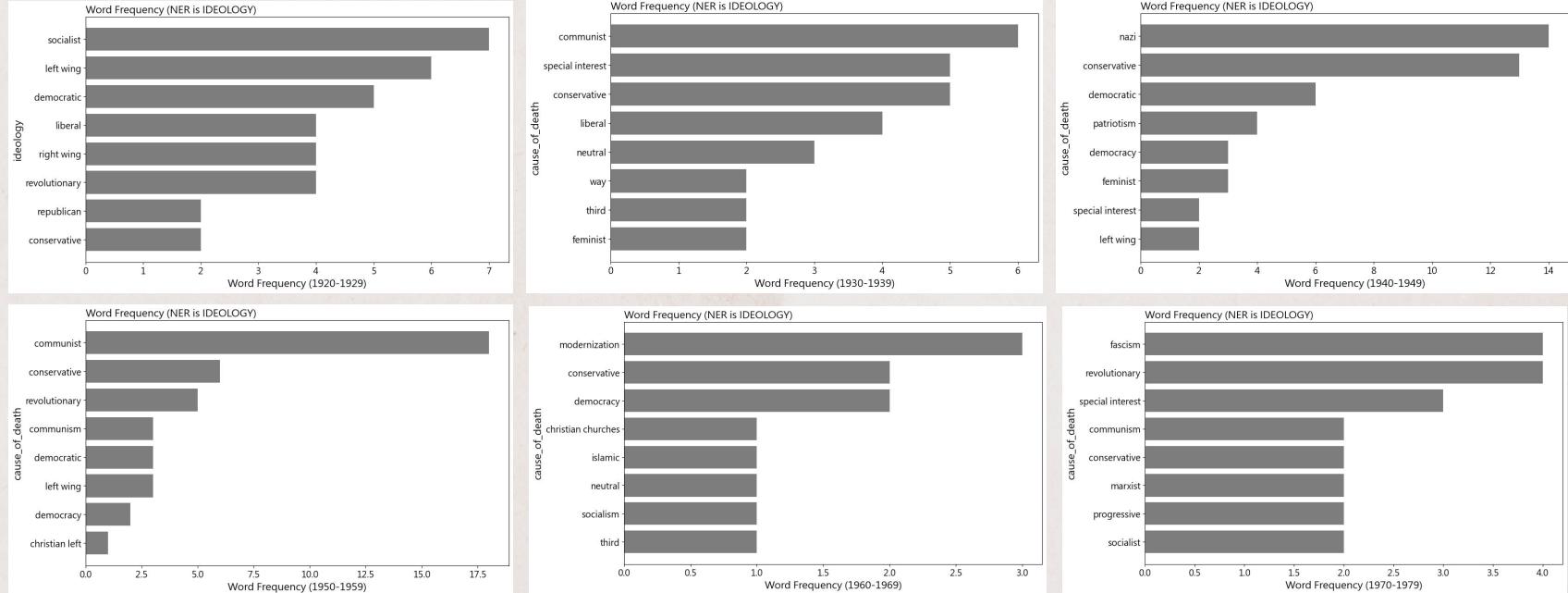
其他詞彙大多圍繞在 **robbery**、**blackmail**、**drugs**、**theft**、**kidnapping** 等，推理小說中常見的犯罪類型。

03 Topic Exploration: NER_Cause of Death



1920s、1940s、1950s，其 war 出現次數較為頻繁，可能與當時社會背景或是戰後氛圍有關。幾乎每個年代都有出現 accident、suicide、poison、fire，等推理小說經典死因。「drugs」、「poison」、「cyanide(氰化物)」顯示出 Agatha Christie 對於各式藥物熟悉。

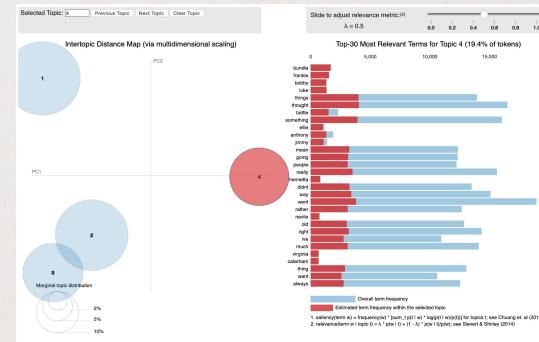
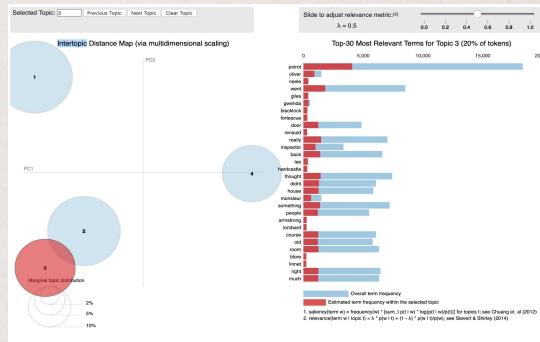
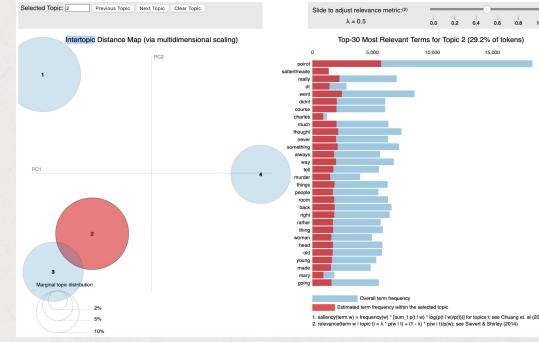
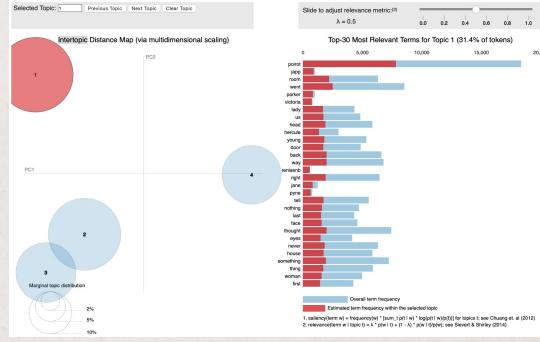
03 Topic Exploration: NER_Ideology



Ideology 可以部份反映到當時的社會背景：

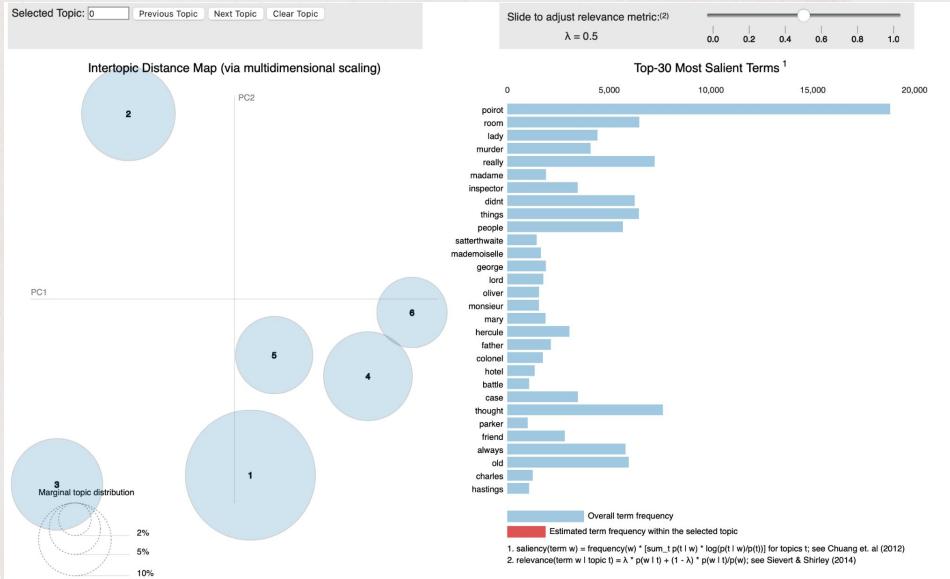
1940s「nazi」顯示二戰時的納粹政權、1950s「communist」顯示冷戰時期共產主義的擴張。
1970s「fascism(法西斯主義)」、「revolutionary」對於戰後社會的不安和焦慮。

03 Topic Exploration: LDA



經過嘗試，LDA在分成4類效果較好，但 無法看出明確分類依據。

03 Topic Exploration: Guided LDA



Seed Word List

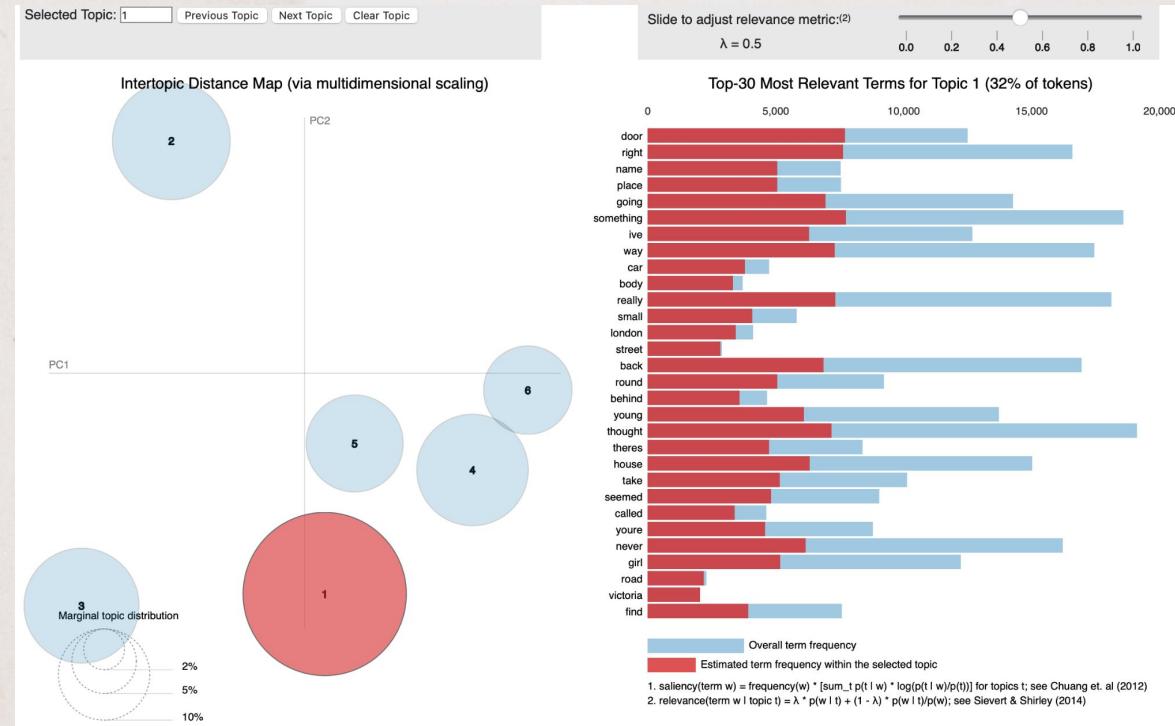
- ['murder', 'poison', 'knife', 'gun', 'strangled', 'crime', 'victim', 'body'] # 殺人事件
- ['detective', 'clue', 'evidence', 'alibi', 'suspicion', 'motive', 'investigation'] # 偵探調查
- ['servant', 'lady', 'maid', 'colonel', 'butler', 'nurse', 'widow', 'inheritance'] # 人物關係 / 社會階級
- ['jealousy', 'greed', 'love', 'hatred', 'revenge', 'fear', 'madness', 'happy', 'afraid', 'scared'] # 情緒
- ['manor', 'village', 'train', 'island', 'library', 'hotel', 'room'] # 地點
- ['deception', 'disguise', 'twist', 'trap', 'secret', 'plan'] # 謊計與反轉

03 Topic Exploration: Guided LDA

Guided LDA 雖然有順利將主題分出，但若細看關鍵字，可以發現仍有一些瑕疵，如 主題1的詞彙似乎並無法明確對應到主題。

由於LDA與GUIDED LDA皆未表現得非常好，我們推測是因為小說是敘事導向，所有 角色、場景等都混雜交錯在長篇文字裡，因此LDA較難準確分類。

改進方式：之後可以再對文本進行更精細的清理，或是 使用單一文本進行分析，而非將全部搜集的文本進行訓練。



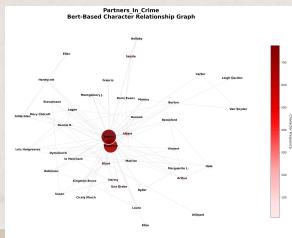
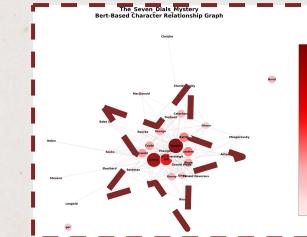
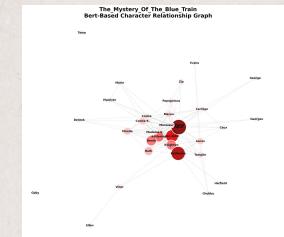
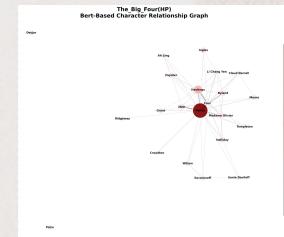
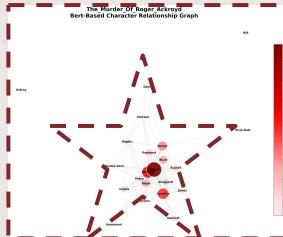
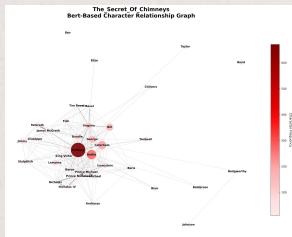
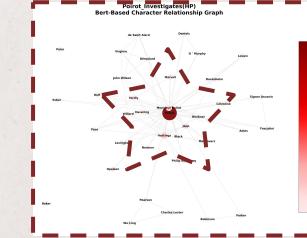
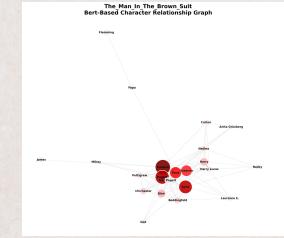
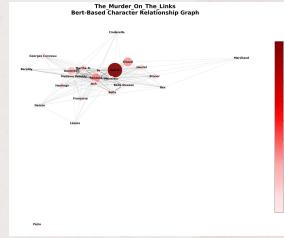
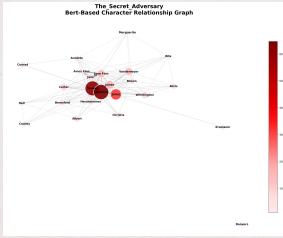
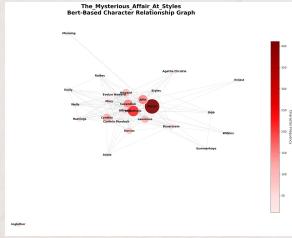
04

Bert-Based
Character Graph

04 Bert-Based Character Graph

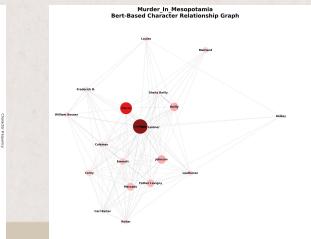
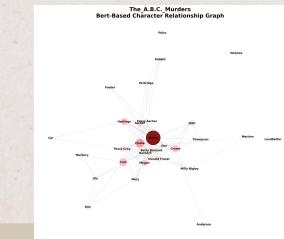
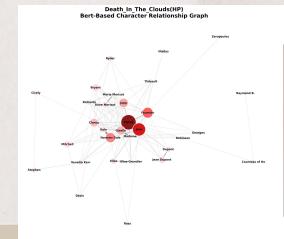
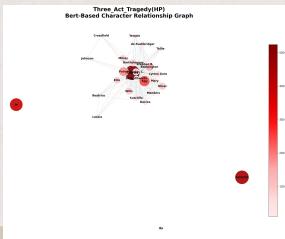
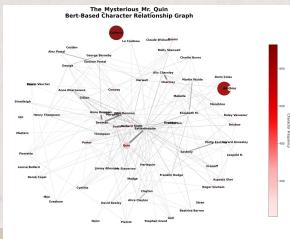
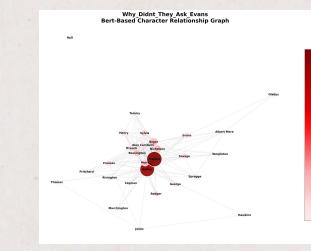
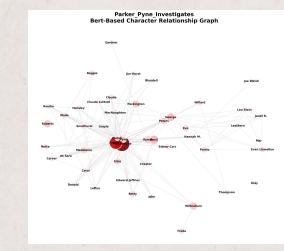
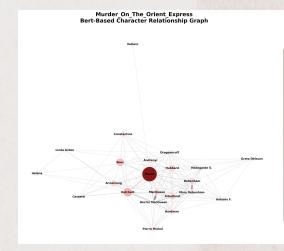
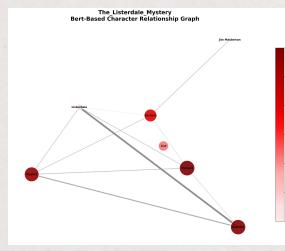
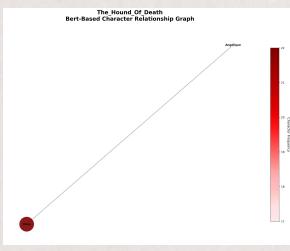
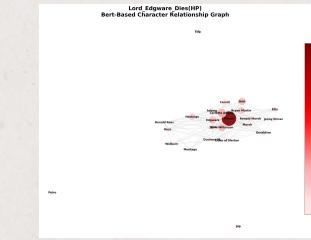
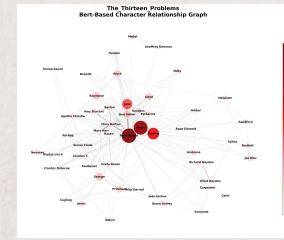
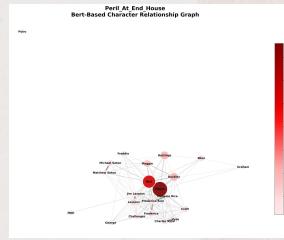
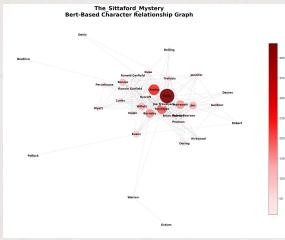
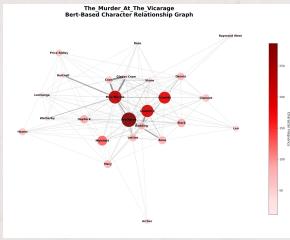
- 使用 BERT NER 模型精準提取小說中的人物名稱
- 結合 Sentence Transformer 識別角色變體
- 透過語義聚類合併角色的不同稱呼
 - e.g. Hercule Poirot = Poirot
- 設定門檻：角色出現 ≥ 10 次，關係強度 ≥ 3

04 Character Graph: 1920-1929

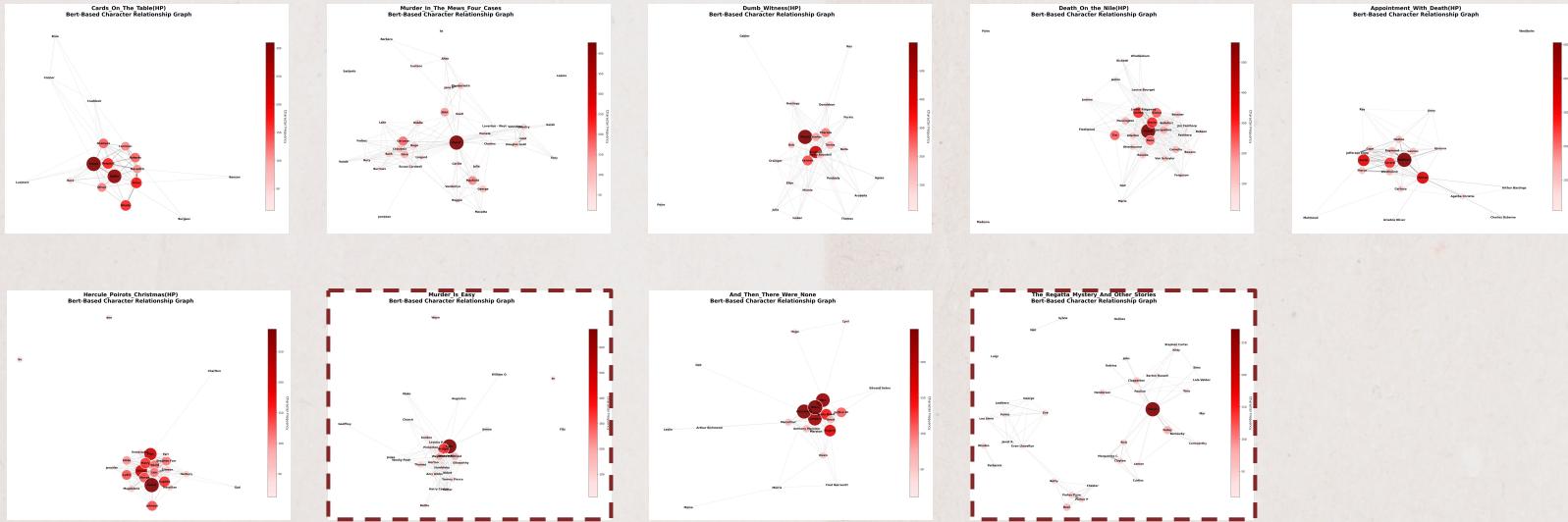


角色關係圖以單一核心角色(如：Hercule Poirot)為中心，故事聚焦於少數關鍵人物的互動，次要角色聯繫較弱，網絡呈星狀結構。

04 Character Graph: 1930-1939 (1)

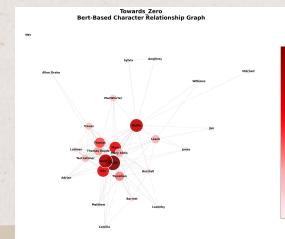
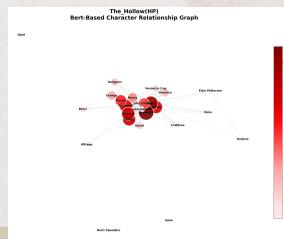
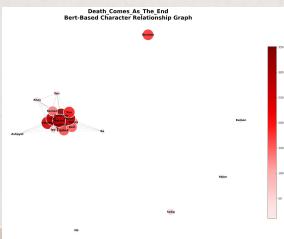
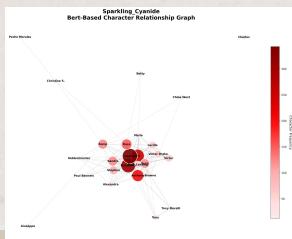
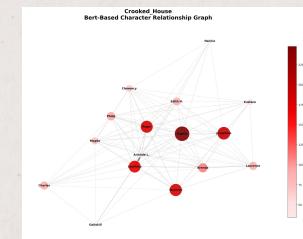
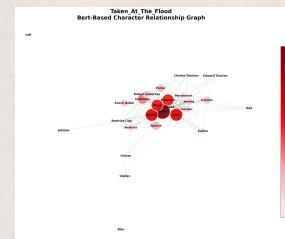
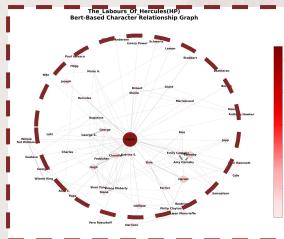
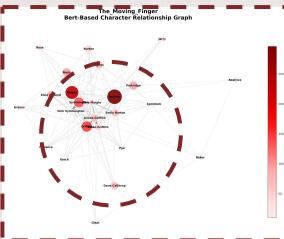
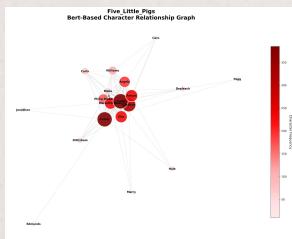
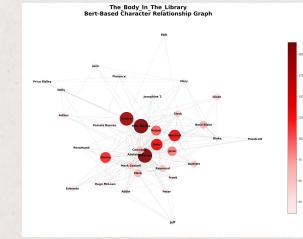
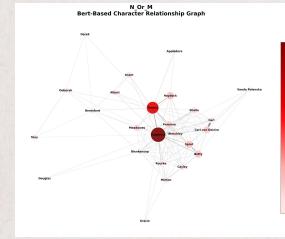
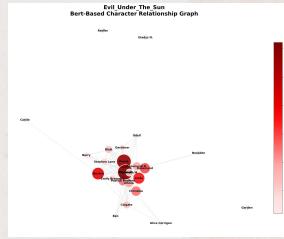
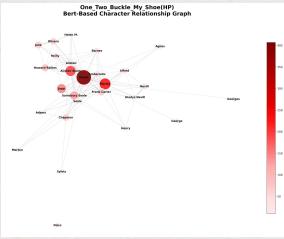
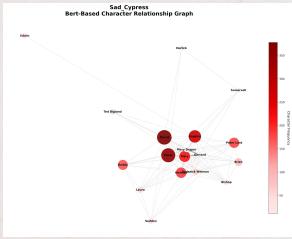


04 Character Graph: 1930-1939 (2)



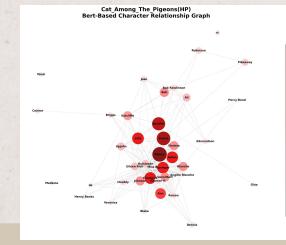
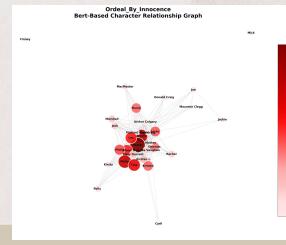
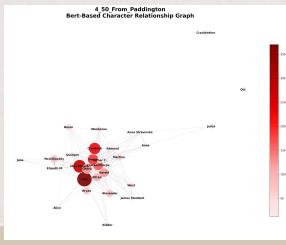
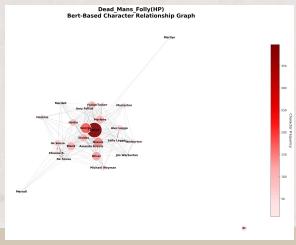
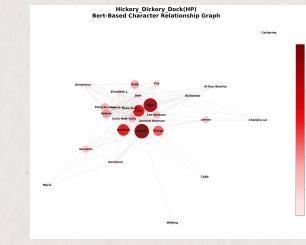
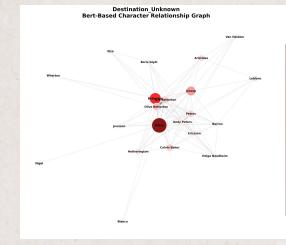
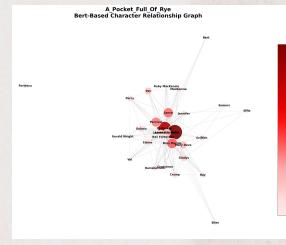
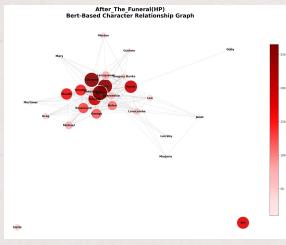
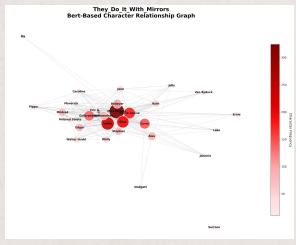
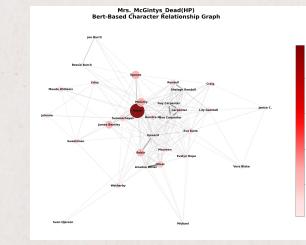
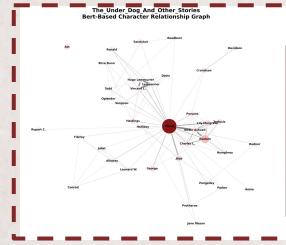
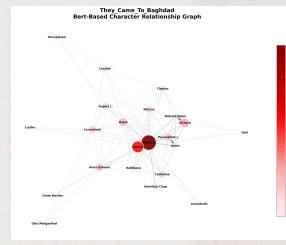
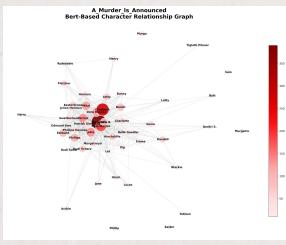
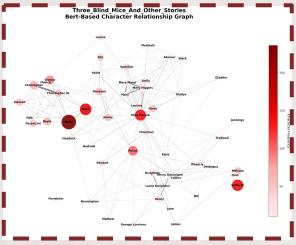
引入 Miss Marple 系列，角色關係圖開始多元化。核心角色周圍出現更多次要角色，網絡結構從 星狀轉向網狀結構，但仍以村莊或小團體為背景，關係較為緊密。

04 Character Graph: 1940-1949



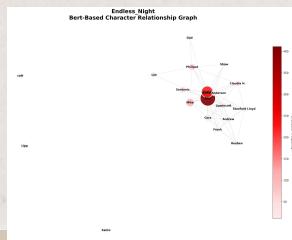
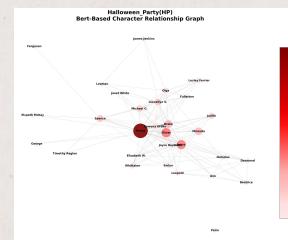
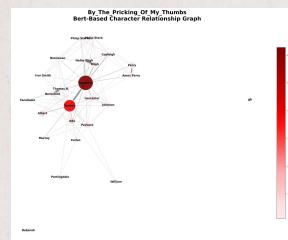
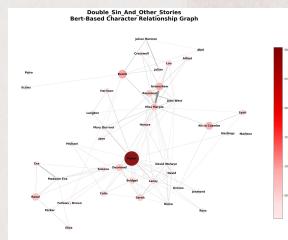
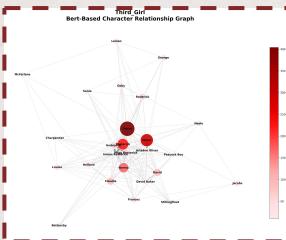
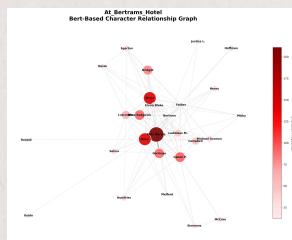
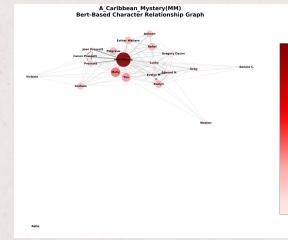
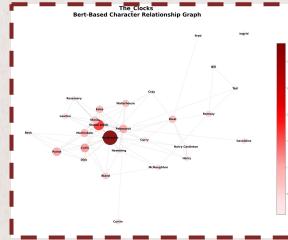
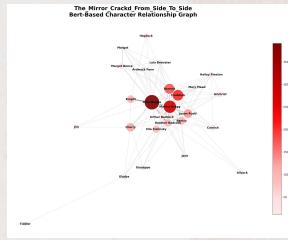
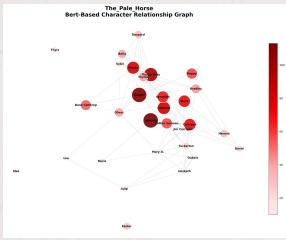
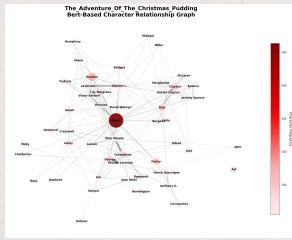
戰時作品角色關係圖可能縮減，聚焦於孤立環境，結構多呈現封閉環狀。

04 Character Graph: 1950-1959



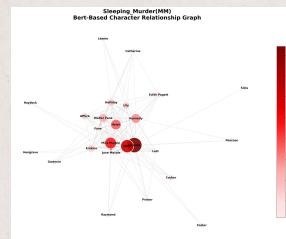
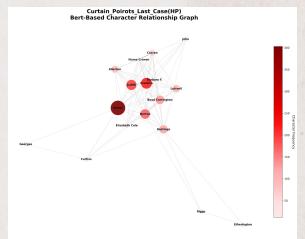
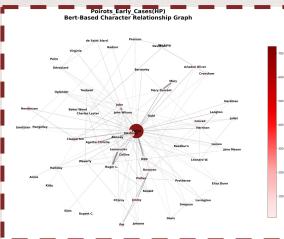
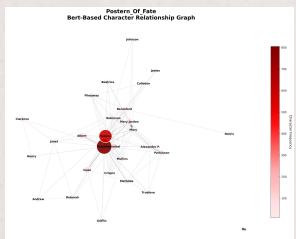
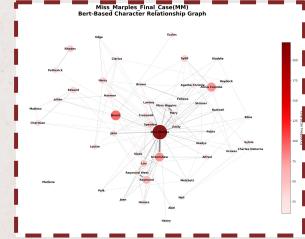
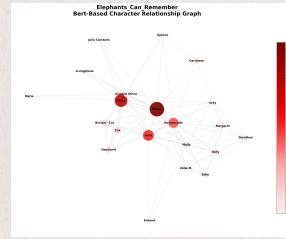
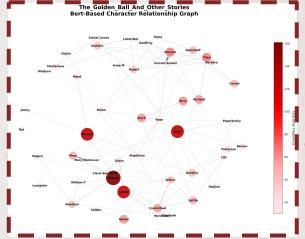
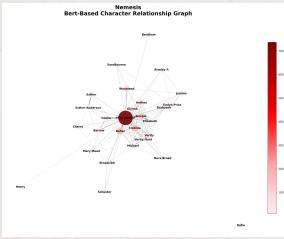
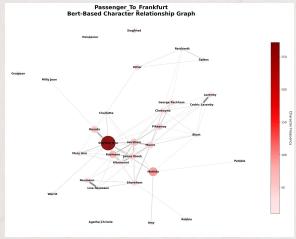
這一時期角色
數量開始增加，呈
分散化結構，顯示
多線敘事特徵。

04 Character Graph: 1960-1969



核心角色(如 :Hercule Poirot 或 Miss Marple)周圍環繞著多樣化次要角色(如:調查對象、目擊者或家庭成員), 角色關係圖繼續呈現 分散化架構。

04 Character Graph: 1970-1979



晚期作品以**角色終結**為主題，關係圖可能回歸簡化，但情感聯繫更深厚。核心角色頻率降低，網狀結構可能代表**情感網絡**而非純偵探網絡，頻率分佈可能更低且均勻。

04 Character Graph: 1920-1979

- **角色數量**: 早期作品 角色數量少，關係集中；後期作品 角色數量多，關係更分散，關係網由星狀向網狀發展。
- **核心角色影響力**: 早期核心角色頻率高、主導性強；後期 核心角色頻率降低、影響力相對分散。
- **關係緊密度**: 早期角色間聯繫緊密，聚焦少數關鍵互動；後期角色間聯繩鬆散，多線敘事特徵明顯。
- **創作風格演變**: 早期集中化結構反映 單一主角或小團體故事；後期分散化結構顯示 多角色交織且劇情更複雜。

Agatha Christie Project: Summary

- **Sentiment Analysis** : 1920-1929 年情感得分顯示較 中性基調，1940-1949 年情感 波動較大 並趨向 負面，1970-1979 年情感 回歸穩定。
- **Topic Exploration** : NER 分析揭示刑事指控相關詞彙頻率高，顯示 犯罪 主題為核心，死亡原因和意識形態分析表明特定時期 社會議題影響顯著，LDA 則需要進行更多處理以近一步分析。
- **Character Graph** : 作品的角色關係圖從 簡單星狀結構，演變為複雜網狀結構互動，敘事中群像描寫逐漸加強。
- **專案延伸** : 開發以阿加莎·克里斯蒂作品為數據庫的 ChatBot，支援回答作品相關問題，並模擬與作者互動，提升讀者沉浸式體驗。目前由於 Google Colab 算力限制無法開發成功。

Thank You
For Listening!

影片連結 : <https://youtu.be/ptYA6FDB0QQ>