

Mapping Eileen Chang's Novels with a Computational Analysis of Themes and Emotions

RUIKAI YUAN

*Faculty of Social Sciences & Humanities
Universiti Kebangsaan Malaysia, Malaysia*

RAVICHANDRAN VENGADASAMY *

*Faculty of Social Sciences & Humanities
Universiti Kebangsaan Malaysia, Malaysia
ravicv@ukm.edu.my*

YU ZHENG

The University of Hong Kong, Hong Kong

ABSTRACT

In the domain of digital humanities, the integration of computational methods with literary studies has opened new approaches to understanding and interpreting literature. This study applies Latent Dirichlet Allocation (LDA) topic modelling and sentiment analysis to eight English novels by Chinese-American author Eileen Chang. The research uncovers seven main themes, reflecting Chang's focus on women's lives, cultural intersections, and societal changes in 1940s Shanghai and Hong Kong. The sentiment analysis of Chang's novels revealed Joy as the dominant positive emotion (19.43%) and Fear and Surprise as the most prominent negative emotion (17.21% each), challenging traditional interpretations of her work as predominantly melancholic. The study demonstrates both the potential and limitations of computational methods in literary analysis. While Natural Language Processing (NLP) techniques provide quantitative support and new perspectives, they sometimes fall short in capturing nuanced literary devices like irony and metaphor. The research advocates the convergence of the traditional literary approach and the burgeoning field of digital humanities to provide a comprehensive perspective.

Keywords: NLP; Eileen Chang; topic modelling; sentiment analysis

INTRODUCTION

Literary analysis has long been the domain of subjective interpretation, where scholars have relied on their insights to understand the narrative, theme, and emotion (Long & So, 2016). In recent years, Natural Language Processing (NLP) has emerged as a transformative tool in the realm of literary studies (Balyan et al., 2017). Its advantages show its ability to dissect, analyse, and represent vast textual data, offering scholars an unprecedented means of the intricate layers of narrative, theme, and style. Digital studies provide new ways to handle this vast amount of information. These new methods give different kinds of proof, which can be mixed with traditional ways of studying topics from more detailed perspectives (Wilkens, 2015).

While NLP has been broadly applied to analyse Western literary corpora (Dai, 2020; Ehrett et al., 2024; Reborá, 2023), its in-depth application to study works of non-Western authors, especially those with the cultural depth of Eileen Chang's writings, remains relatively unexplored (Yuan et al., 2024). Eileen Chang, also known as Zhang Ailing, is a renowned Chinese-American writer whose works have captivated readers for decades (Louie, 2012). She was a well-known feminist woman writer of Chinese literature, known for portraying life in 1940s Shanghai and

Hong Kong. Born in Shanghai in 1920, Chang's novels dissected personal relationships set against the backdrop of significant historical events, such as the Second Sino-Japanese War. Chang's work is highly respected and has had a lasting impact on Chinese literature. Hu et al. (2022) believe that her work is "intensely introspective" and "filled with fragility of life". Her works are renowned for their powerful insights and innovative use of language. Chang's novels and short stories focus on the lives of women in pre-Communist China and explore themes of love, identity, and human relationships (Ren, 2020). The selection of Chang's literary works for this research is motivated not only by her significant literary stature but also by her combination of Eastern and Western cultural components found in her novels, which renders them highly suitable for a detailed investigation utilising NLP techniques.

The objectives of this study are to use NLP to investigate the following issues: (1) conducting topic modelling to identify and analyse dominant themes and motifs across Chang's works, (2) undertaking sentiment analysis (SA) to discern the emotional tones and differences embedded within Chang's narratives, (3) explore the discrepancy of traditional qualitative analysis and computational method in studying literary works. As the boundaries between traditional literary scholarship and digital humanities are less distinct, this research stands as a testament to the confluence of these two domains, aiming to offer a comprehensive perspective on Chang's works.

LITERATURE REVIEW

NLP IN LITERATURE

Natural Language Processing (NLP), a sub-field of artificial intelligence, has made significant strides in recent years. While traditionally associated with linguistics and computer science, NLP's applications have extended to the realm of literary studies, offering novel methodologies to analyse and interpret literary texts. The integration of computational methods in literary analysis is not entirely new. Franco Moretti's work on "distant reading" in the early 2000s underscored the potential of using computational methods to analyse vast literary corpora (Moretti, 2013). However, with the advent of sophisticated NLP techniques, the depth and breadth of such analyses have expanded dramatically. Jockers and Mimno (2013) pioneered the use of topic modelling on a large corpus of 19th-century novels, uncovering patterns related to plot development. Their study found that external elements like the writer's gender, nationality, and when a novel was published can influence its themes and how they are expressed, using statistics to find themes from 19th-century novels from Britain, Ireland, and the US applications of digital methods in literary studies. Underwood (2019) explores how digital tools and methods, particularly computational approaches, can offer new insights into literary evolution and transformations over extended periods. These automated analyses have illuminated previously undetected patterns in thematic evolution. The application of quantitative methods to humanities, long resisted by traditionalists, has witnessed an uptake due to improvements in computer science and natural language processing. Andrew Piper (2019) investigates the fundamental facets of literature, from poetic punctuation to novel plots, striving to bridge the divide between numeric data and literary essence. As part of the burgeoning digital humanities movement, Piper's approach reflects a broader academic trend that seeks to harmoniously blend traditional literary critique with data-driven analyses, marking a transformative moment in literary studies.

The transformative integration of NLP techniques into literary analysis has predominantly focused on Western literary corpora, leaving a significant research gap regarding non-Western literature (Stranisci et al., 2022). This study's primary objective is to bridge this gap by applying advanced NLP techniques, such as topic modelling and SA, to the works of Chang, a revered Chinese-born American author renowned for her unique narrative style and intricate portrayals of life in 1940s Shanghai and Hong Kong. This study aims to merge Chang's culturally rich narratives with computational methodologies, highlighting the versatility of NLP in analysing diverse literary texts and fostering a deeper, more integrative understanding of non-Western literary masterpieces in the realm of digital humanities.

LDA TOPIC MODELING

Topic modelling is a computational method that identifies latent topics within a large text collection. In essence, the topic modelling algorithm categorises groups of documents into coherent themes without requiring user input. Topic modelling is considered a clustering issue that involves a random combination of themes inside a document, with each word assumed to be associated with at least one of those topics (Boon-Itt & Skunkan, 2020; Mutanga & Abayomi, 2022). Several topic modelling approaches have been used in the literature, including Latent Dirichlet Allocation (LDA), Document-pivot topic detection (Doc-p) (Winarko & Pulungan, 2019), Frequent Pattern Mining (FPM), Graph-based feature-pivot topic detection (GFeat-p), Soft Frequent Pattern Mining (SFPM) (Zendah & Maghari, 2019), and the BNgram approach (Ibrahim et al., 2018).

The most widely recognised method for topic modelling is LDA (Albalawi et al., 2020; Gerlach et al., 2018). This strategy utilises a "bag of words" methodology, where each text is treated as a vector of word counts. LDA has gained popularity in text retrieval applications and has demonstrated outstanding performance in other domains, such as entity resolution systems, fraud detection in telecommunication systems, and picture processing. Mohammed and AI-augby (2020) conducted a comparison study on e-books, demonstrating that LDA outperformed Latent Semantic Analysis (LSA) in terms of coherence value for text document classification. Similarly, LDA demonstrated higher accuracy compared to the LSI method in terms of topic modelling, indexing and visualisation, particularly excelling in word indexing for topics from tweet data (Negara et al., 2019). LDA was also used to identify the veracity of news (Casillo et al., 2021; Elhadad et al., 2020). The use of LDA for topic modelling in literature research has also been explored in the context of digital humanities, revealing latent patterns and divergences in literary texts (Broadwell et al., 2019). This method has been used to analyse themes of literary texts if it is applied to a certain quantity of appropriately cleaned text (Uglanova et al., 2020).

SENTIMENT ANALYSIS

Sentiment analysis, or opinion mining, refers to the computer information technology capable of computational analysis of text viewpoints, emotions, and subjectivity. As an essential branch within NLP, SA can accomplish relevant NLP tasks at different textual levels (Asif et al., 2020; Stine, 2019). Recent decades have seen rapid growth in the use of SA. It has been applied in different areas to detect writer's opinions in fiction, such as Harry Potter (Jacobs, 2019), Lovecraft (Arroyo-Barrigüete, 2023), George Orwell (Dunder & Pavlovski, 2019) and the plays of Shakespeare (Nalisnick & Baird, 2013). Eileen Chang was famous for her exquisite writing skills

in expressing the psychology of characters. It is clear that her writing style shows obvious differences in her life phrases; her writing was usually distinguished by three phases: the early period, pre- and post-immigration to America (Chen, 2020). Despite that, scholars have a consensus that there was a transition point in Chang's writing career (Hoyan, 1996); that is, after she immigrated to America, Chang's later novels were interpreted in many ways.

METHOD

This section presents the research methods utilised in this study. We collected e-books by Eileen Chang as raw data from an open-access online resource, Library Genesis. The raw data were cleaned and processed using the Natural Language Toolkit (NLTK) in Python and were inputted in text format. The process involves removing stop words (e.g. "a," "the," "is," and "are"), named entities like people and place names, and irrelevant terms. Then, the study employed LDA for topic modeling and NRC Word-Emotion Association Lexicon for SA in Python. The research approach was divided into three sections, as described in Figure 1, including Data Collection, Pre-processing, and Data Analysis.

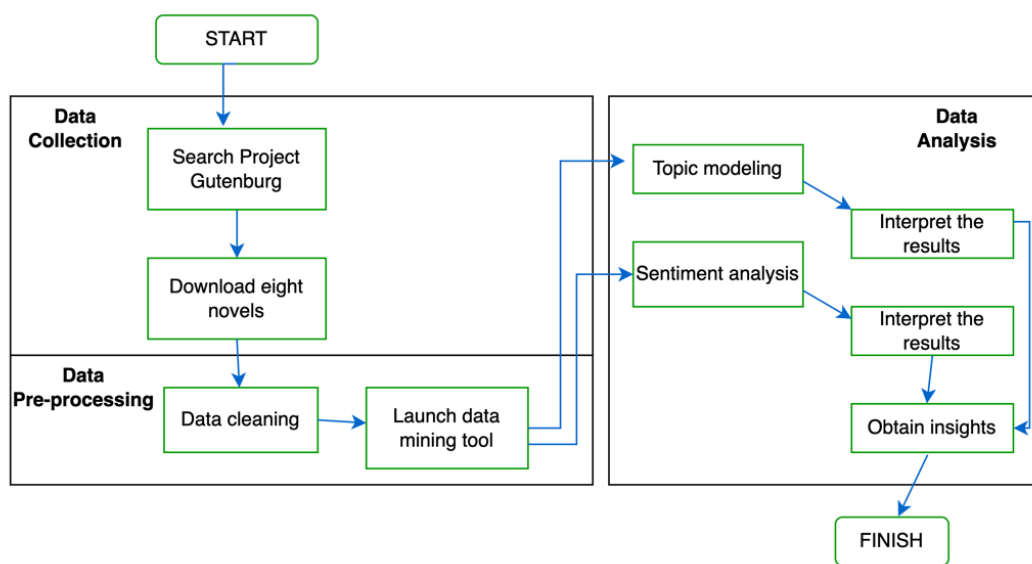


FIGURE1. Research Approach

DATA COLLECTION

This study included a selection of eight English novels by Eileen Chang (Table 1). After pre-processing, the corpus contains 13,862 tokens. The corpus-based method can assist subsequent topic analysis (Taufek et al., 2022). The English version was used for this study because our primary objective is not to examine the discourse or cultural context between translated books and their Chinese counterparts. Besides, Chang's international renown and influence in the world of literature further justify our choice to focus on the English versions. This approach allows us to explore her impact on a global literary scale.

TABLE 1. Summary for each novel within this study

<i>No.</i>	<i>Title</i>	<i>Year of Publication</i>
1	<i>Love in a Fallen City</i>	1943
2	<i>Traces of Love</i>	1945
3	<i>Half a Lifetime Romance</i>	1951
4	<i>The Rice Sprout Song</i>	1952
5	<i>Naked Earth</i>	1954
6	<i>The Rouge of the North</i>	1967
7	<i>Little Reunions</i>	1975
8	<i>Lust, Caution</i>	1978

DATA PRE-PROCESSING

The eight novels were pre-processed by breaking the texts into small pieces of 200 words for each paragraph as it was considered representative of each topic (Chiru et al., 2014). Then, the texts were tokenised into words, phrases, or bi-grams called tokens. In data pre-processing, it is essential to eliminate irrelevant words from the text, often referred to as stop-words. Common stop words such as "of", "and", and "them" were typically removed as they do not contribute significantly to the thematic content of the text. We employed Matthew Jocker's expanded stop-word list for 19th-century British, American, and Irish fiction was often used, which includes thousands of personal names common in the 19th century, to remove the stopwords often seen in literary works in datasets. The remaining tokens are generally stemmed, which reduces inflected words to their base or root form. For instance, the words "specialised" and "specialisation" might be reduced to the morphological root "specialise." We also converted all vocabulary to lowercase to ensure uniformity and consistency in our dataset. Lastly, tokens were lemmatised to convert verbs to their infinite tense.

DATA ANALYSIS

TOPIC MODELLING

The foundational idea of LDA is that every document encompasses multiple themes, each with its word likelihood distribution (Toutanova & Johnson, 2007). Based on this distribution, a topic from the distribution of topics is randomly chosen for each word of the document (Blei, 2012). For example, in a collection of novels, a topic related to war might include words such as *battle*, *soldier*, *victory*, and *enemy*, while a topic associated with romance might feature words like *heart*, *passion*, *love*, and *longing*. By analysing the frequency and co-occurrence of words across texts, LDA reveals these themes and quantifies their presence in each document.

A useful analogy to explain LDA is imagining a library without labelled sections. If one were to scan the books and group them based on recurring keywords, shelves might naturally emerge: books containing words like *dragon*, *kingdom*, and *sword* might be grouped under the fantasy genre, while others with terms like *detective*, *crime*, and *murder* could belong to the mystery category. LDA functions similarly, identifying clusters of words that tend to occur together and treating each cluster as a distinct "topic." However, unlike traditional genre classification, LDA recognises that a single document can contain multiple topics. For instance, a novel might predominantly discuss themes of war but also include passages focused on romance.

LDA assigns proportional values to each topic, showing the extent to which they contribute to the overall content of the text.

At the mathematical level, the process of LDA can be modelled by a three-level Bayesian graphical model, where random variables are represented by nodes and possible dependencies between the variables are represented by edges, as depicted in Figure 2. In this representation, α refers to the Dirichlet parameter, Θ refers to document-level topic variables, z refers to per-word topic assignment, w refers to the observed word, and β refers to the topics. As can be observed from the three-layered representation, α and β parameters are sampled once while generating the corpus, document-level topic variables are sampled for each document, and word-level variables are sampled for each word of the document.

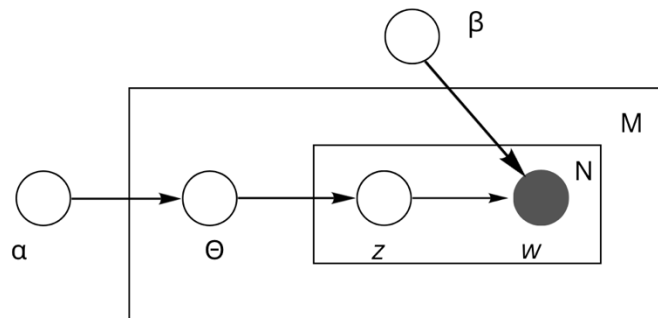


FIGURE 2. The graphical representation of LDA

The generative process of LDA indicates a joint distribution over random variables. The probability density function of a k -dimensional Dirichlet random variable is computed by Equation 1, the joint distribution of a topic mixture is determined by Equation 2, and the probability of a corpus is computed as given by Equation 3 (Blei et al., 2003):

$$p(\theta | \alpha) = \frac{r(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k r(\alpha_i)} \theta_1^{\alpha_1-1} \dots \theta_k^{\alpha_k-1} \quad (1)$$

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta) \quad (2)$$

$$p(D | \alpha, \beta) = \prod_{n=1}^M \int p(\theta_d | \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d \quad (3)$$

The pre-set number of topics, k , profoundly influences the output of the LDA model. The choice of this number is guided by the research's purpose and desired outcomes (Boussalis & Coan, 2016; Quinn et al., 2010). Setting k too high might yield fragmented results, while a value too low might lead to excessive grouping (Greene et al., 2014). In this research, LDA was executed with themes set at 7, 10, and 15 to assess the quality of outcomes for an optimal k value. The results were found satisfactory when $k=7$, which was deemed fitting for the research. The LDA modelling was executed in Python using the Gensim library (Rehurek & Sojka, 2010).

SENTIMENT ANALYSIS

Sentiment analysis encompasses two predominant methodologies: wordlist-based sentiment analysis, wherein words are assigned scores to determine sentiment and machine learning-based SA (Stine, 2019). The machine learning-based method requires substantial labelled training data to achieve accuracy, which was not feasible given the scope and resources of this study. Therefore, we used the wordlist-based methodology, specifically the NRC Word-Emotion Association Lexicon, which is often used to monitor the emotional states expressed by individuals and the distribution of these emotions. The NRC Word-Emotion Association Lexicon (Mohammad & Turney, 2013) is frequently employed to derive emotional content from textual data. This lexicon comprises a comprehensive compilation of 10,170 English words, each assigned codes corresponding to Plutchik's eight fundamental human emotions (namely anger, fear, anticipation, surprise, trust, joy, sadness, and disgust), as well as two polarities, namely positive and negative. The scoring system assigns a numerical value between 0 and 1 to each word, indicating the range of emotional intensity associated with that particular term. This process was conducted in Python using NLTK.

RESULTS AND DISCUSSION

The following part begins with the results of topic modelling by LDA. Then, the results of the sentiment analysis are presented in terms of eight different emotions that are distributed across Chang's novels.

TOPIC MODELING AND ANALYSIS

A total of seven distinct topics have been identified in the study. Each topic is characterised by a collection of ten keywords that effectively encapsulate the central themes associated with that specific topic. Arranged in descending order based on their frequency, the list of topics is as follows: Domestic Life, Political Undertones, Cultural Fusion, Traditional Symbols, Generation Transition, Marriage and Family Struggles, and Opulence and Hong Kong. Figure 3 displays the proportional distribution of each topic within the novels.

(1) Domestic Life

The first theme encompasses everyday routines, relationships, familial affairs, and the intersection of personal spaces, both literal and metaphorical. The keywords associated with this topic, such as “house”, “child”, “money”, “girl”, “money”, “shanghai”, and “light”, not only indicate familial and household affairs but also suggest the socio-economic challenges that intertwine with them.

“Domestic Life” is consistently present in all of Chang’s works, implying its significance in her narratives. Particularly prominent in *Half a Lifelong Romance* and *The Rouge of the North*, the theme captures the essence of the daily life and struggles of ordinary individuals against the broader canvas of societal changes. In a society where women’s lives were often predetermined by societal norms, the domestic realm became both a sanctuary and a battleground. The mention of “money” in the context of domestic life also hints at the economic dependencies and constraints that women often face, further complicating their roles and choices (Zheng, 2017). The prominence of this theme in her pre-1955 works, set primarily in Shanghai and Hong Kong, offers

an insight into the domestic scope of these cities during a period of tumultuous change. Chang's fine portrayal of domestic life goes beyond mere depiction; it becomes a lens through which readers can discern the larger socio-cultural and political transformations unfolding in China during that era.

(2) Cultural Fusion

The next topic, "Cultural Fusion" in Chang's narratives, emphasises the blending of Eastern and Western cultural elements. This theme, with key terms like 'Shanghai', 'piano', and 'party', is emblematic of the intertwining of two worlds. One of the most prominent manifestations of this fusion is through music, notably the piano – a symbol of Western classical music, finding its place in the traditionally Eastern setting of her stories.

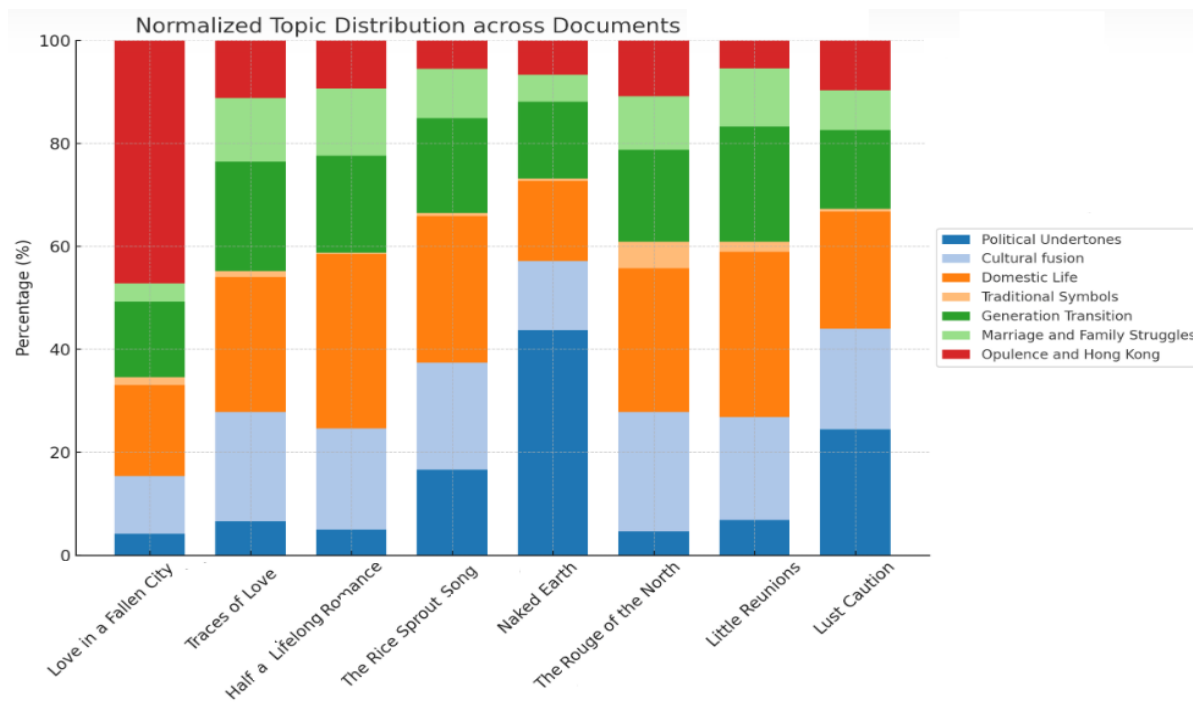


FIGURE 3. Thematic Distribution across Chang's Novels

Shanghai, during the early 20th century, was partitioned into various foreign concessions. This made it a unique epicentre of global influences, especially Western. On the other hand, Hong Kong was under British colonial rule and was deeply influenced by British culture. The influence of the British garden party, an epitome of Western social gatherings, is evident in Chang's depiction of "parties". Such parties became venues where the East met the West, where traditional Chinese music might meld with Western tunes, and where Chinese elites and foreigners might intermingle.

Cultural fusion in works *Love in a Fallen City*, and *The Rouge of the North* reflects the settings of cosmopolitan cities like Shanghai and Hong Kong. In *Love in a Fallen City*, In these cities, the amalgamation of Chinese traditions with Western modernity was not just physical but also psychological. Chang's characters often find themselves at a crossroads, torn between traditional values and the allure of the West. As stated by C. Yang (2016), Chang's story usually takes place in Shanghai and Hong Kong; the different perspectives of story space narration in the

novel, the different values and positions expressed in the space, the different characters' different perspectives in the space, and the different meanings of the space. The physical space of the novel is mainly about the urban space of Shanghai and Hong Kong, while the psychological space makes full use of metaphors and symbols to express her inner feelings.

(3) Political Undertones

“Political Undertones” is a recurring theme throughout Chang’s works, including the socio-political atmosphere of China during the mid-20th century, with keywords such as “comrade”, “office”, “man”, “land”, and “turn”. The term “comrade”, a common salutation among communists, indicates the change in social relationships in the wake of the Communist Party’s rise to power. The words in Chang’s works underline the pervasive influence of politics in everyday life. Moreover, the inclusion of words like “office” suggests a focus on bureaucracy, land reforms, and the struggles of the common man during these transformative years.

Eileen Chang captures the nuanced shifts in societal structures, personal relationships, and individual aspirations against this political backdrop. Her works, set in this era, are not just mere stories of love and loss but are also intricate tapestries that intertwine personal narratives with the everchanged political environment. The prominence of this theme in *Naked Earth* and *The Rice Sprout Song* stands out, which in particular looks into the lives of individuals during the land reforms and the rise of communism, making the “Political Undertones” theme especially poignant. These works were completed in the middle and late stages of Chang’s life. According to X. Li (2022), after 1949, under the change of literary policy, Chang turned to make her literary creations dependent on different political forces, which led to her creations showing a lack of an objective narrative. Through this theme, Chang offers readers a lens to view the upheaval of life during one of China’s disorderly periods.

(4) Traditional Symbols

This theme is rich with keywords that evoke deep cultural and historical sentiments associated with China. Words such as ‘twin’, “simplicity”, “benevolence”, “phoenix”, “prosperity”, “jade”, “treasure”, “lotus”, “blossom”, and “pearl” are not just mere words; they encompass centuries of Chinese tradition, beliefs, and values. The phoenix, for instance, is a mythological bird in Chinese culture representing high virtue and grace, symbolising the union of yin and yang. Similarly, jade is not only a precious stone but is also associated with purity, nobility, and immortality in Chinese culture.

The “Traditional Symbols” theme is not the dominant theme in any of the novels. However, it’s subtly present in all of them, suggesting that while Chang delved into the complexities of modern relationships and societal changes, she never completely detached her narratives from the traditional backdrop. This could be indicative of the transitional phase that Chinese society was undergoing at the time – moving from traditionalism to modernism, but still deeply rooted in its past. Most of Chang’s novels focus on the subject of marriage, showing the dark side of the old-fashioned big families in the city (Hu, 2023). For instance, the presence of this theme in *Love in a Fallen City* suggests how, even in the rapidly changing landscape of post-war Hong Kong, traditions and symbols continued to play an essential role in the lives of individuals. Similarly, in *The Rice Sprout Song*, set against the backdrop of the Chinese Civil War, the subtle presence of traditional symbols could be emblematic of the society’s struggle to hold onto its identity amidst political upheaval.

(5) Generation Transition

The theme “Generation Transition” shows the trifle between different age groups, capturing the shifts in beliefs, values, and aspirations across time. Keywords such as “child”, “young”, “old”, “master”, and “mistress” hint at the relationships and contrasts between the young and the old, potentially pointing to generational conflicts and changes in societal norms. The terms “dinner”, “money”, “table”, and “party” suggest social settings where these interactions and contrasts might be most palpable.

“Generation Transition” is a recurring theme in Chang’s works but is most pronounced in *Traces of Love* and *Half a Lifelong Romance*. In these two novels, Chang’s works embody a remarkable intergenerational continuity and intergenerational rupture due to her influence by the complexity of her family’s interpersonal relationships (Huang, 2022). For example, Julie and her mother in *Half a Lifelong Romance* similarly cried over watching the film, and in *Traces of Love*, Chang wrote three times about Chunyu Dunfeng’s influence by the aunts in her family of origin. The artistic expression of Chang’s intergenerational cyclopean narrative is extremely layered, which helps to portray characters, unearth hidden psychology, deepen the theme of the text, and highlight Chang’s view of social development.

(6) Marriage and Family Struggles

“Marriage and Family Struggles” is a continuous theme in Chang’s oeuvre, capturing interpersonal relationships, social expectations, and personal aspirations. This theme is characterised by keywords such as “manservant”, “marriage”, “marry”, “piano”, “jewel”, “rouge”, “party”, “rustic”, “house” and “father”, each echoing various facets of familial ties and societal norms around matrimony.

This theme shares a nearly similar weight in all the novels. Chang reveals the real state of women’s existence in social life. Whether it is the perverted and hostile Cao Qiqiao, Bai Liusu or Nixi, all these women are struggling to obtain or maintain a legal marriage in their lives because marriage is the only way out of their survival.

(7) Opulence and Hong Kong

The topic “Opulence and Hong Kong” is characterised by keywords such as “car”, “diamond”, “ring”, “moment”, “dinner”, “cape”, “road”, and “black”. This theme seems to be intricately tied to the glitz, glamour, and rapid transformation of Hong Kong as a bustling metropolis in the mid-20th century.

“Opulence and Hong Kong” is most prominently featured in *Love in a Fallen City*. This suggests that the narrative of this work revolves heavily around the urbane, luxury, and the fast-paced world of Hong Kong. The opulence represents not just the physical luxury but also the metaphorical richness of the city’s cultural tapestry during the time. The usage of terms like “car”, “diamond”, and “ring” indicate a certain class distinction, wealth, and the aspiration of individuals living in the city. On the one hand, it signifies prosperity and growth, but on the other, it can imply disparity, superficiality, and the darker aspects of urban life. Hong Kong, post-World War II, was transitioning into a financial hub, attracting people from different walks of life. The detailed portrayal, on the one hand, can reflect her attitude towards traditional civilisation and modern urban civilisation; on the other hand, it also reflects her complex and contradictory psychology towards the discourse of enlightenment and modernity (X. Li, 2011).

Looking more closely at the thematic distribution in Chang's novels, it is evident that there are two prominent features: exploring the conflict between tradition and modernity (mainly related to topics 2 "Cultural Fusion" and 5 "Generation Transition"), and depicting the details and struggles of everyday life (mainly related to topics 1 "Domestic Life" and 6 "Marriage and Family Struggles").

Regarding the first feature, many of Chang's works deal with the collision between tradition and modernity while also mentioning characters who struggle in this collision. This is probably due to the backdrop of rapid social change in mid-20th century China. Some examples include the mention of "In Hong Kong, we follow the British way." (Chang, 2006, p. 45) The piano theme, which forms contrasts with traditional Chinese culture: "Although Weilong "took lessons for a few years," she "is not very good" (41) at the instrument." (41) The contrast between "young" and "old", suggesting conflicts in values between different generations. As for depicting the details and struggles of everyday life, the data shows that this seemingly serves as Chang's way of illustrating broader social contexts. Some examples of this are:- as long as you live, you belong to your family, and once you die, your spirit will belong to them too," is one that is widely held and largely accepted across society (2006: 113). The mention of "house", "child", and "money" in the "Domestic Life" theme, reflecting the daily life and economic pressures of ordinary people: "Why was I such a fool? All because of that fuss over money, little bits of money, now I have no way out of here!" (Chang, 2006: 89). Words like "manservant", "marriage", and "marry" reveal the complexity of marital and family relationships in the society of that time: "[George is] born to be married into a wealthy family." (Chang, 2006:67)

According to the results, what Chang cares most about in the collective sense is the impact of cultural collisions and social changes, then the survival state and actions of individuals under this background. This reflects Chang's profound insight into her era and her writing technique of reflecting social changes through individual destinies. The seven topics reveal the main concerns in Chang's novels about Chinese society in the mid-20th century, as well as her in-depth exploration of character destinies.

SENTIMENT ANALYSIS

The sentiment analysis was conducted in Python with the NRC Word-Emotion Association Lexicon. This analysis allowed for the examination of the distribution and relative representation of the eight emotions as defined by Plutchik within the dataset. Plutchik (1980, 1982) introduced the concept of the wheel of emotions, which organises emotions into eight distinct categories: joy, trust, fear, surprise, sadness, disgust, rage, and anticipation. This theory has had a significant influence on the field of psychology. The individual proposed a conceptual framework wherein the eight emotions exhibit a bipolar structure. Specifically, joy is posited as the antithesis of sadness, fear is juxtaposed with wrath, anticipation is contrasted with surprise, and disgust is positioned as the opposite of trust. Joy, trust, anticipation, and rage are commonly seen as emotions with a positive valence, indicating a favourable or pleasant state. Conversely, fear, sadness, disgust, and surprise are associated with a negative valence, indicating an unfavourable or unpleasant state. Contrary to prevailing beliefs, rage is categorised as a positive feeling despite its negative connotation, while surprise is regarded as a negative emotion. The reason can be attributed to the perception of anger as a manifestation of strength, serving as a catalyst for action (Hess, 2014) and reflecting a trajectory of progress towards a desired objective (TenHouten, 2014). The concept of

surprise encompasses an infringement against individuals' psychological boundaries and typically elicits negative emotions (Noordewier & Breugelmans, 2013).

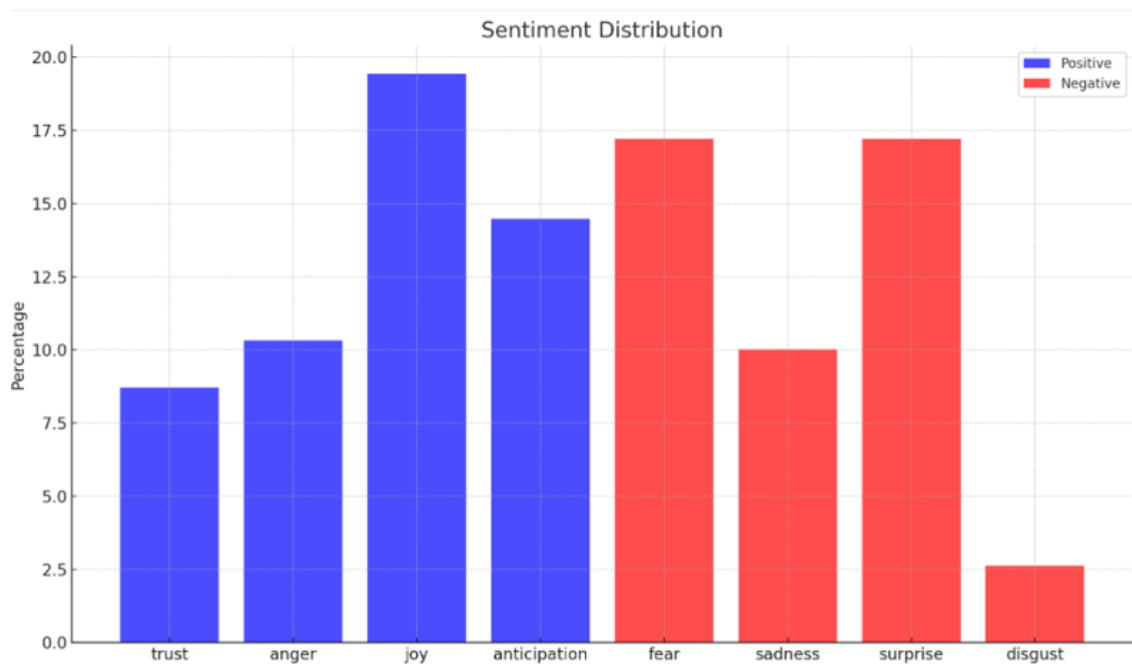


FIGURE 4. Emotion proportions

Chang's novels exhibit a range of emotional subtleties that can be categorised into positive and negative affective dimensions (Figure 4). The dominant sentiment within the positive affective domain is Joy, accounting for 19.43% of the recognised sentiments. The subsequent element in the sequence is denoted as Anticipation, constituting a proportion of 14.47%. Interestingly, the emotion of Anger, which is commonly classified as a negative effect in traditional sentiment analysis, is observed to account for 10.32% of the data. The percentage of Trust is situated at the lower echelon, specifically at 8.70%. In contrast, within the context of negative affect, both Fear and Surprise exhibit equal prominence, with each accounting for 17.21% of the total. The emotion of Sadness is observed to have a prevalence of 10.02%, but the emotion of Disgust is found to be little prominent, accounting for only 2.63% of the total. When considering both positive and negative emotional indicators together, it becomes apparent that there is a slightly greater presence of positive emotional tones, particularly with Joy being the most prominent. However, the prominent occurrence of Fear and Surprise within the negative category highlights a complex emotional fabric that is intricately threaded throughout the novels.

To conduct a more comprehensive analysis of Chang's books, this study employed the Groupby sum method to explore the relationship between emotions and the narrative. The result is presented in Figure 5.

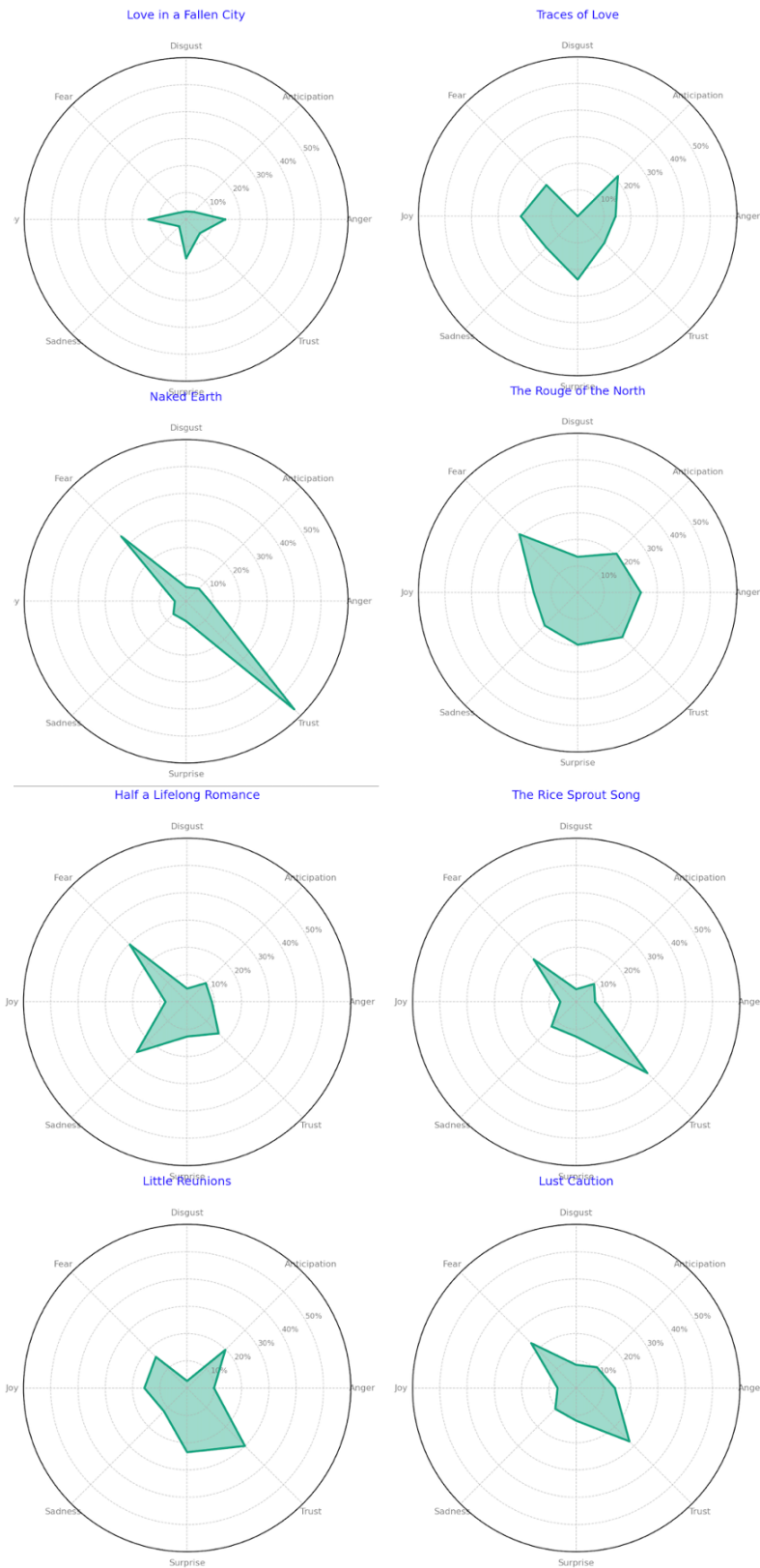


FIGURE 5. Emotion distribution in each novel

In the emotional analysis of novels, Chang offers a compelling insight into the thematic and tonal shifts in her work over the years. *Love in a Fallen City* (1943) stands out with a striking blend of Anger and Surprise, suggesting a narrative filled with unexpected twists and underlying tensions. Interestingly, the novel lacks indications of Anticipation, Disgust, or Joy, perhaps pointing to a story devoid of hopeful prospects or sheer revulsion. It is very worth noticing that *Traces of Love* (1945) is rich in a variety of emotions, with high levels of Anticipation, Joy, Sadness, and Surprise, painting a picture of a narrative that is both uplifting and heart-wrenching. The absence of Disgust once again hints at the absence of elements that provoke sheer aversion. In *Half a Lifelong Romance* (1951), Fear emerges as the dominant emotion, suggesting themes of tension, anxiety, or conflict in the storyline, which is further complemented by notable representations of Sadness and Trust. *The Rice Sprout Song* (1952) centres on Trust, which might indicate a narrative anchored in relationships, alliances, or faith. This emphasis on Trust continues in *Naked Earth* (1954), though it is juxtaposed with a dominant tone of Fear, perhaps pointing to a story of trust amidst threatening circumstances. *The Rouge of the North* (1967) offers a balanced emotional palette, with a slight tilt towards Anger, Fear, and Sadness, which could imply a multifaceted storyline with varying emotional highs and lows. *Little Reunions* (1975) leans heavily on Trust, further enriched by elements of Sadness, Surprise, and Anticipation. This mix might indicate a storyline that revolves around relationships, perhaps with unexpected turns and hopeful undertones. Lastly, *Lust, Caution* (1978) is coloured by Fear, suggesting a narrative filled with tension or potential danger but also enriched by elements of Trust and Surprise.

The SA results show different results compared to the previous understanding of Chang's emotions in her writings. The conventional academic analysis of *Little Reunions* has emphasised its emotional inclination toward Gothic stylisations and unconventional psychological themes (Chen, 2020; Han, 2015). The Gothic genre, known for its distinct dark atmospherics, exhibits a connection to emotional data by incorporating Fear and Sadness (Snodgrass, 2014). However, the analysis falls short of accurately capturing the profound and all-encompassing nature of these feelings, which is often expected in the Gothic genre. Psychological conditions proposed by traditional research appear to be more complex and sophisticated than the simple alternation between Trust and Surprise indicated by the emotional data. Within the domain of romantic representation, conventional perspectives provide a trace in which Chang's depiction of love progresses from profound anguish to a comparatively restrained and almost tacit harmony with nature in her subsequent literary creations. The emotional data, on the other hand, has similarities to this trace through its expression of Trust and Sadness. However, it does not possess the same level of profound desolation and realistic undertones that are frequently emphasised in conventional interpretations. The inclusion of Joy, however, to a limited extent, in the emotional evaluation presents a slight contradiction to the subdued portrayals of love proposed by conventional interpretations. The emotional data reflects the sad and lonely picture of Chang's female heroines, as seen in Chen's research (Chen, 2020). Nevertheless, the significant emphasis placed on Trust in the emotional analysis appears to present a more sanguine or positive connotation compared to conventional interpretations. While there is a shared understanding between both techniques about the melancholic nature of these characters, the inclusion of emotional data offers a dimension of resilience or a need for comfort that may not completely fit with the deep sense of despair put forth by conventional interpretations.

CONCLUSION

This study employed natural language processing techniques, specifically LDA topic modelling and sentiment analysis, to analyse Eileen Chang's eight novels in English.

Topic modelling uncovered seven main themes in Chang's works: domestic life, political undertones, cultural fusion, traditional symbols, generation transition, marriage and family struggles, and opulence in Hong Kong. The most prominent theme, domestic life, was consistently present across all of Chang's works, particularly in *Half a Lifelong Romance* and *The Rouge of the North*. This theme captures the essence of daily life and the struggles of ordinary individuals against the canvas of social/ changes. The emergence of the "political undertones" theme, especially prominent in *Naked Earth* and *The Rice Sprout Song*, challenges the view that Chang's works are "depoliticised", revealing the significance of political elements in her writing.

The sentiment analysis results provide a more nuanced understanding of the emotional landscape in Chang's novels. Contrary to traditional interpretations of her work as predominantly melancholic, the analysis revealed Joy as the dominant positive emotion (19.43%) and Fear and Surprise as the most prominent negative emotion (17.21% each). This finding offers a counterpoint to conventional readings of Chang's work. Interestingly, the emotion distribution varied across different novels. This variation in emotional content across her works suggests an evolution in Chang's writing style and thematic focus over time.

This research contributes to the growing field of digital humanities by offering a data-driven approach to analysing Chang's novels, complementing traditional literary criticism. It demonstrates both the potential and limitations of computational methods in literary analysis, suggesting the potential for applying similar methodologies to other authors' works and exploring the evolution of themes and sentiments across different literary periods or cultural contexts. While NLP techniques provide quantitative support and new perspectives, they sometimes fall short in capturing literary devices like irony and metaphor. Future studies could address these shortcomings by integrating more advanced techniques, such as context-aware models like transformer-based architectures (e.g., BERT or GPT), which can better understand the subtleties of language through contextual embeddings.

REFERENCES

- Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). 'Using topic modelling methods for short-text data: A comparative analysis'. *Frontiers in Artificial Intelligence*, 3(42), pp.1–14. doi:10.3389/frai.2020.00042 PMID:33733159
- Arroyo-Barrigüete, J. L. (2023). Sentiment analysis of Lovecraft's fiction writings. *Heliyon*, 9(1).
- Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H., & Shah, J. (2020). 'Sentiment analysis of extremism in social media from textual information'. *Telematics and Informatics*, 48, 1–20. doi:10.1016/j.tele.2020.101345
- Balyan, R., McCarthy, K. S., & McNamara, D. S. (2017). Combining Machine Learning and Natural Language Processing to Assess Literary Text Comprehension. *Grantee Submission*.
- Blei, D. M. (2012). 'Introduction to probabilistic topic models'. *Communications of the ACM*, 55(4), pp.77–84. doi:10.1145/2133806.2133826
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). 'Latent dirichlet allocation'. *Journal of Machine Learning Research*, 3, 993-1022.
- Boon-Itt, S., & Skunkan, Y. (2020). Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. *JMIR public health and surveillance*, 6(4), e21978.
- Boussalis, C., & Coan, T. G. (2016). 'Text-mining the signals of climate change doubt'. *Global Environmental Change*, 36, 89–100. doi:10.1016/j.gloenvcha.2015.12.001
- Broadwell, P., Chen, J. W., & Shepard, D. (2019). Reading the Quan Tang shi: Literary History, Topic Modeling, Divergence Measures. *Digital Humanities Quarterly*, 13(4).

- Casillo, M., Colace, F., Gupta, B. B., Santaniello, D., & Valentino, C. (2021). Fake news detection using LDA topic modelling and K-nearest neighbor classifier. In *Computational Data and Social Networks: 10th International Conference, CSoNet 2021, Virtual Event, November 15–17, 2021, Proceedings 10* (pp. 330-339). Springer International Publishing.
- Chang, E. (2006). *Love in a fallen city*. New York Review of Books.
- Chen, C. (2020). On the "Gothic style" of "Little Reunion". [Dissertation]. Xiamen University. (Chinese)
- Chiru, C., Rebedea, T., & Ciotec, S. (2014, April). Comparison between LSA-LDA-lexical chains. In *International Conference on Web Information Systems and Technologies* (Vol. 2, pp. 255-262). SCITEPRESS.
- Dai, Y. (2020). *A New Literary Analysis of The Great Gatsby by Natural Language Processing* (Doctoral dissertation, Doshisha University).
- Dunder, I., & Pavlovski, M. (2019, May). Behind the dystopian sentiment: a sentiment analysis of George Orwell's 1984. In *2019, 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 577-582. IEEE.
- Ehrett, C., Ghita, L., Ranwala, D., & Menezes, A. (2024). Shakespeare Machine: New AI-Based Technologies for Textual Analysis. *Digital Scholarship in the Humanities*, 39(2), 522-531.
- Elhadad, M. K., Li, K. F., & Gebali, F. (2020). A novel approach for selecting hybrid features from online news textual metadata for fake news detection. In *Advances on P2P, Parallel, Grid, Cloud and Internet Computing: Proceedings of the 14th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2019) 14* (pp. 914-925). Springer International Publishing.
- Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances*, 4(7), pp.1–11. doi:10.1126/sciadv.aag1360 PMID:30035215
- Greene, D., O'Callaghan, D., & Cunningham, P. (2014). *How many topics? Stability analysis for topic models*. 10.1007/978-3-662-44848-9_32
- Han, J. (2015). A study on Eileen Chang's novels from the perspective of abnormal psychology. [Dissertation]. Northeast Normal University.
- Hess, U. (2014). 'Anger is a positive emotion', in *The positive side of negative emotions*, pp. 55–75. The Guilford Press.
- Hoyan, C. H. (1996). *The life and works of Zhang Ailing: a critical study* (Doctoral dissertation, University of British Columbia).
- Hu, X. (2023). 'A Brief Analysis of Eileen Chang's Creative Style, Female Consciousness, and Female Care.' *Literature and Art Weekly*, 2, pp.49-52.
- Hu, J. (2022). The collision between drama and film of a love legend: a review of the Tim Yip version of Love in a Fallen City. *Drama Home*, 36, 6-8.
- Huang, Q. (2022). 'On the Generational Spiral Narrative of Eileen Chang's Works.' *Chinese Modern Literature Research Series*, 3, 213-227. doi:10.16287/j.cnki.cn11-2589/i.2022.03.004.
- Ibrahim, R., Elbagoury, A., Kamel, M. S., & Karray, F. (2018). Tools and approaches for topic detection from Twitter streams: survey. *Knowledge and Information Systems*, 54, 511-539.
- Jacobs, A. M. (2019). 'Sentiment analysis for words and fiction characters from the perspective of computational (neuro-) poetics'. *Frontiers in Robotics and AI*, 6, 53-54.
- Jockers, M. L., & Mimno, D. (2013). 'Significant themes in 19th-century literature'. *Poetics*, 41(6), 750-769.
- Li, X. (2011). 'Discussing Modern Urban Civilization in Eileen Chang's Eyes: Taking *Love in a Fallen City* as an Example.' *Lanzhou Academic Journal*, 6, 172-185.
- Li, X. (2022). *'On the Gains and Losses of Eileen Chang's Creation from the Perspective of Political Culture'*. Hebei University.
- Long, H., & So, R. J. (2016). 'Literary pattern recognition: Modernism between close reading and machine learning'. *Critical inquiry*, 42(2), 235-267.
- Louie, K. (Ed.). (2012). *Eileen Chang: Romancing Languages, Cultures and Genres* (Vol. 1). Hong Kong University Press.
- Mohammad, S. M., & Turney, P. D. (2013). 'Crowdsourcing a word-emotion association lexicon'. *Computational Intelligence*, 29(3), 436–465. doi:10.1111/j.1467-8640.2012.00460.x
- Mohammed, S. H., & Al-augby, S. (2020). Lsa & lda topic modeling classification: Comparison study on e-books. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(1), 353-362.
- Moretti, F. (2013). *Distant reading*. Verso Books.
- Mutanga, M. B., & Abayomi, A. (2022). Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. *African Journal of Science, Technology, Innovation and Development*, 14(1), 163-172.

- Nalisnick, E. T., & Baird, H. S. (2013, August). Character-to-character sentiment analysis in Shakespeare's plays. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 479-483.
- Negara, E. S., Triadi, D., & Andryani, R. (2019, October). Topic modelling twitter data with latent dirichlet allocation method. In *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)* (pp. 386-390). IEEE.
- Noordewier, M. K., & Breugelmans, S. M. (2013). 'On the valence of surprise', *Cognition and Emotion*, 27(7), pp. 1326–1334. <https://doi.org/10.1080/02699931.2013.777660>
- Piper, A. (2019). *Enumerations: data and literary study*. University of Chicago Press.
- Plutchik, R. (1980). 'A general psychoevolutionary theory of emotion', in *Theories of emotion*, pp. 3–33. Elsevier. <https://doi.org/10.1016/B978-0-12-558701-3.50007-7>
- Plutchik, R. (1982). 'A psychoevolutionary theory of emotions', *Social Sciences Information*. Information Sur les Sciences Sociales, 21(4–5), pp. 529–553. <https://doi.org/10.1177/053901882021004003>
- Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., Radev, D. R., Purpura, S., Schrod, P., Sin, G., Sinclair, B., Ward, M., Wilkerson, J., Wood, D., & Zorn, C. (2010). 'How to analyse political attention with minimal assumptions and costs'. *American Journal of Political Science*, 54(1), 209–228. doi:10.1111/j.1540-5907.2009.00427.x
- Rebora, S. (2023). Sentiment Analysis in Literary Studies. A Critical Survey. *DHQ: Digital Humanities Quarterly*, 17(3).
- Rehurek, R., & Sojka, P. (2010). 'Software framework for topic modelling with large corpora'. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Doi:10.1.1.695.4595
- Snodgrass, M. E. (2014). *Encyclopedia of Gothic literature*. Infobase Publishing.
- Stine, R. A. (2019). Sentiment analysis. *Annual Review of Statistics and Its Application*, 6(1), 287–308. doi:10.1146/annurev-statistics-030718-105242
- Stranisci, M. A., Spillo, G., Musto, C., Patti, V., & Damiano, R. (2022). The URW-KG: a resource for tackling the underrepresentation of non-western writers. *arXiv preprint arXiv:2212.13104*.
- Taufek, T. E., Nor, N. F. M., Jaludin, A., Tiun, S., Choy, L. K., & Aziz, J. (2022). Media Report About Climate Change in an English Online Malaysian Newspaper Through Thematic and Discourse Analysis Approaches. *3L: Southeast Asian Journal of English Language Studies*, 28(2).
- TenHouten, W. D. (2014). *Emotion and reason: Mind, brain, and the social domains of work and love*. Routledge. <https://doi.org/10.4324/9780203093634>
- Toutanova, K., & Johnson, M. (2007). 'A Bayesian LDA-based model for semi-supervised part-of-speech tagging'. *Advances in neural information processing systems*, 20.
- Uglanova, I., Gius, E., Karsdorp, F., McGillivray, B., Nerghes, A., & Wevers, M. (2020). The Order of Things. A Study on Topic Modelling of Literary Texts. *CHR*, (18-20), 2020.
- Underwood, T. (2019). *Distant horizons: digital evidence and literary change*. University of Chicago Press.
- Wilkens, M. (2015). 'Digital humanities and its application in the study of literature and culture'. *Comparative Literature*, 67(1), 11-20.
- Winarko, E., & Pulungan, R. (2019). Trending topics detection of Indonesian tweets using BN-grams and Doc-p. *Journal of King Saud University-Computer and Information Sciences*, 31(2), 266-274.
- Yang, C. (2016). 'Narrative Space in Eileen Chang's Novels.' *Academic Exploration*, 6, 135-141.
- Yuan, R., Vengadasamy, R., & Zheng, Y. (2024). A Bibliometric Analysis of Study on Eileen Chang Using CiteSpace. *SAGE Open*, 14(2).
- Zheng, S. (2017). 'A Comparison of Family Views in Maugham and Eileen Chang's Novels.' *Journal of Nanjing University of Science and Technology (Social Science Edition)*, 2, 67-71.
- Zendah, J. H., & Maghari, A. Y. (2019). Detecting Significant Events in Arabic Microblogs using Soft Frequent Pattern Mining. *Journal of Engineering Research & Technology*, 6(1).