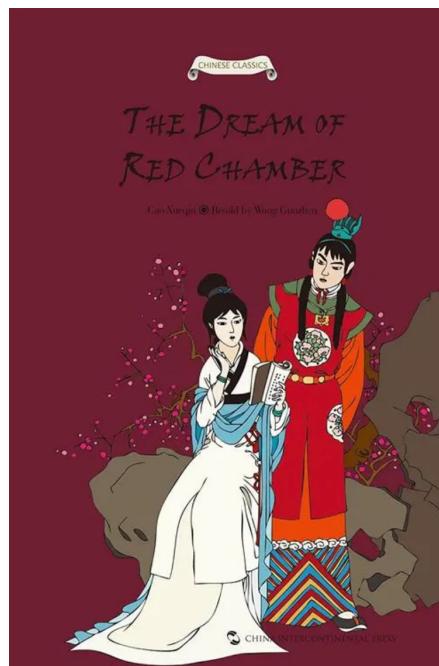


Discourse Analysis of Lin Daiyu Compared to Other Characters in THE DREAM OF THE RED CHAMBER



Github : <https://github.com/Harper0927/TOOLS-FinalProject-GroupF.git>

Group Project Report

Yaran Lu(S5482488), Mengqi Li(S5479606),

Shuwei Wang(S5466121), Zixuan Wu(S5483662),

Yuxin Huang(S5574242), Jiaming Chen(S5692302)

Faculty of arts, University of Groningen

LHU011M05: Digital Humanities: Tools and Methods

Dr. Federico Pianzola Dr. Lucas van der Deijl

January 17, 2024

Table of Contents

Introduction and Motivation	3
Tools and Methods	6
BookNLP and Python	9
Voyant	12
Word Frequency	12
Word Collocation	18
Deeper Analysis Combined with Voyant	25
Limitations of Tools	31
BookBLP	31
Voyant	34
Conclusion	36
References	39

INTRODUCTION AND MOTIVATION

THE DREAM OF THE RED CHAMBER was initially called *THE STORY OF STONE* and was written by Cao Xueqin during the Qianlong period of the Qing Dynasty. It is a chapter-style classical novel belonging to one of the four great Chinese masterpieces. The novel portrays the ascent and decline of the four prominent families - Jia, Shi, Wang, and Xue unfolding against the backdrop of the intricate upper-class aristocratic society of the Qing Dynasty to provide a vivid depiction of China's feudal society in the first half of the 18th century. Throughout the novel, the love tragedy between Jia Baoyu and Lin Daiyu serves as its central theme, and the narrative captures the complexity of that era, thus portraying the rich tapestry of human emotions within the worldly setting.

The author, Cao Xueqin, was born into a wealthy family similar to the Jia family depicted in the article. However, owing to the dynasty shifts, the Cao family eventually became a declining aristocrat. He lived a wealthy life when he was young but gradually fell into poverty as a teenager. Cao was versatile and good at poetry and painting, dedicating his entire life to creating this literary masterpiece. Cao Xueqin's experience of significant changes from prosperity to poverty gave him a clear-eyed view of society and human nature. He integrated his rebellious spirit into the book and decisively criticized the feudal orthodox thoughts, expressing his longing for a humanistic society. Baoyu and Lin Daiyu are the two characters who embody his preliminary democratic ideas of freedom of thought, liberation of individuality, and equal human rights(Zhu,1994). There is an acknowledged comment to describe his

commitment to the book: "Every word seems to be filled with blood. His ten-year time devotion is precious and unusual." (The Dream of the Red Chamber Collector's Edition, 2014)

This project aims to focus on digital tools and quantitative research approaches, which are learned this semester on text analysis. Thus, the popularity and literary value of the text are essential measures of the work chosen for this project. *THE DREAM OF THE RED CHAMBER* is an excellent work in both aspects. Firstly, it is one of the four great masterpieces of China, which has been translated into many languages, including English, Dutch, French, Spanish, and Russian, showing this novel's global influence and literary value. After its abridged and full versions in different languages were published in the 19th century, *THE DREAM OF THE RED CHAMBER became the* “most widely read Chinese novel in the Western world” (Poupard & Dennis, 1984). The author tailors each character's imagery and language to their social status, personality traits, and life experiences. The distinctive language styles of different characters allow digital tools to work better, thereby helping quantitative analysis methods to explore the differences in this language style.

Moreover, this novel has also attracted the attention of digital humanist Franco Moretti and has been applied to his research. In the article *NETWORK THEORY, PLOT ANALYSIS*, he used this novel as a research subject. When studying how the overall network shapes individuals, he chose Baoyu, the main male character in this novel, as his main research subject. Moretti studied the conversations and network between Baoyu and other characters in the first twenty chapters of *THE DREAM OF*

THE RED CHAMBER and expressed his sympathy for such a protagonist who is constrained and restricted: “constantly summoned by this and that relative, kept under supervision, asked to perform all sorts of duties—even the many delightful opportunities he is offered usually come with constraints attached. The protagonist, yes, but not free”. (Moretti 2011)

However, what he failed to notice was that although Baoyu was seen as a victim of that era, he also enjoyed the preferential treatment and privileges of a feudal family. In contrast, the female characters in the book are more oppressed and show a stronger sense of rebellion. This novel has 36 main female characters: First Register of Twelve Beauties of Jinling, Second Register of Twelve Beauties of Jinling, and Third Register of Twelve Beauties of Jinling. The registers are distinguished by social class; from the first to the third, they respectively present upper-class miss, medium-class concubine, and lower-class servant. The author wrote about their oppressed fates but also gave them their unique charms. This kind of charm transcends the constraints of class and language. Thus, more attention should be paid to female characters in research based on the English translation.

Furthermore, this work was written during the Qing Dynasty and has the attributes of China's feudal society. Such a work that creates a group portrait of women, especially one that treats women as independent individuals and shows their talents and charm instead of being accessories to their husbands, is courageous and rare. Therefore, digital humanities research should also pay more attention to the female characters in this work (Poupard, Dennis, 1984). Consequently, this project

chose one of the most unique female characters in the work as our main subject: Lin Daiyu, who belongs to the First Register of Twelve Beauties of Jinling.

As per the book's narrative, Daiyu lost her mother at a young age. After that, she was subsequently taken in by her maternal grandmother to be raised in the Jia family, sharing a similar raising environment to Baoyu's. Despite struggling with frailty and frequent illnesses since childhood, she exhibited remarkable intelligence and sensitivity. Growing up alongside Baoyu, they developed a profound connection, yet, in the end, they did not become a married couple.

Many sentimental women have been in literary works throughout the ages, but finding a character who can be compared to Lin Daiyu is not easy. She is not only the first heroine in the story, but this character also carries the projection of Chinese national culture. Through such an artistic image, the understanding of beauty in ancient Chinese traditional culture can be seen concretely (Lyu, 1987). Although Lin Daiyu and Jia Baoyu jointly become the author's emotional carriers in this literature, Lin Daiyu's plot development and symbolic meaning differ from those of the other protagonists. In Lin Daiyu's portrayal, the author depicts a woman pursuing an independent personality, seeking personal development, and hating feudal forces and all hypocritical things (Zhu, 1994). In conclusion, Lin Daiyu is the most appropriate research object for this project.

TOOLS AND METHODS

Based on the selection of text and dominant research object, the research question of this project is settled as: How does the language style of discourses, characterized by the specific mentions of Lin Daiyu's name (including Lin Daiyu, Daiyu), differ from the rest of the characters in the Chinese classic *THE DREAM OF THE RED CHAMBER*? There are two reasons for choosing Lin Daiyu and Daiyu as two retrieve names. One is to ensure the accuracy of extracting Lin Daiyu's discourse. This explicit reference guarantees that the selected discourses clearly belong to Lin Daiyu, avoiding the inclusion of other characters' discourses. Another one is to resolve the potential doubts of personal pronouns and ambiguous verbal expressions, refraining from conversations that are difficult to distinguish accurately by machine. Lin Daiyu, as a character, stands out within the intricate novel, and by scrutinizing her language style specifically, this project aims to unravel unique linguistic features and expressive nuances that set her apart from the broader linguistic landscape of the discourse. Ultimately, by intertwining the historical and societal perspectives with the examination of Lin Daiyu's language styles, this research project strives to provide a comprehensive understanding of the cultural and social dynamics shaping the linguistic expressions of Lin Daiyu in *THE DREAM OF THE RED CHAMBER*.

With the purpose of restoring the emotions and language style that the original work wants to convey to the greatest extent possible, the English version of "Dream of the Red Chamber," translated by Xianyi Yang, will be chosen as the metadata throughout the project. Yang's version has been considered one of the most

authoritative and recognized translators by university scholars in China (Benmin,2011). Narration is always mixed with comments, and the subjective feelings of the author are attached to objective descriptions at any time throughout the novel. This is a typical way of expression in Chinese, which is seldom seen in English. Many translators tend to delete the subjective comments under such circumstances, for it is not the usual way of expression in English. Still, the Yangs preserved this kind of structure in his translation and adopted the foreignization strategy to preserve the original Chinese culture as much as possible (Benmin,2011). Therefore, Yang's translation stands out as this project's original data. The first dataset will be applied based on the original data, extracting Lin Daiyu's discourse from the complete novel employing BookNLP. Afterward, the second dataset will first extract all discourse from the book, and the final result is the elimination of Lin Daiyu's discourse from the complete one. Both steps are done by Python. In conclusion, two datasets are used overall to study the research question.

Apart from the programming code employed in the extraction of Lin Daiyu's discourse, A tool learned from this block will be applied to this project as well: Voyant. Voyant, an online text analysis tool that provides simple and intuitive visualizations of the text, is suitable for text exploration, offering insights into critical features and frequency statistics with various graphics. Application of this digital humanities tool to this project will specifically present on Word/Collocation frequencies and text visualization.

BookNLP and Python

As introduced in the previous section, this project expected to obtain the first dataset, Lin Daiyu's discourse, with the help of BookNLP and employed Python to get the second dataset. However, due to the unpredictably wrong recognition and division of characters and quotes done by BookNLP, Python has been applied to process the output further to make the first dataset more accurate and comprehensive. Therefore, this part will not only demonstrate the applications of BookNLP and Python but also present the process of resolving difficulties and the limitations of tools discovered by the practice throughout the project.

BookNLP is a natural language processing (NLP) tool designed for the analysis and extraction of information from literary texts. This project aimed to extract Lin Daiyu's discourses from the overall book employing BookNLP. After operating BookNLP in the Python environment within Anaconda, the comprehensive analysis and the divisions of different dimensions are generated automatically into the designated output directory. Meanwhile, the time demanded to run the code to get each result is listed respectively (Figure 1). Eventually, there are, in total, four text files representing entities, quotes, supersense, tokens, and one HTML file (Figure 2).

```
--- spacy: 76.958 seconds ---
--- entities: 442.519 seconds ---
--- quotes: 0.234 seconds ---
--- attribution: 440.532 seconds ---
--- name coref: 2.988 seconds ---
--- coref: 449.400 seconds ---
--- TOTAL (excl. startup): 1414.165 seconds ---, 769634 words
```

Figure 1

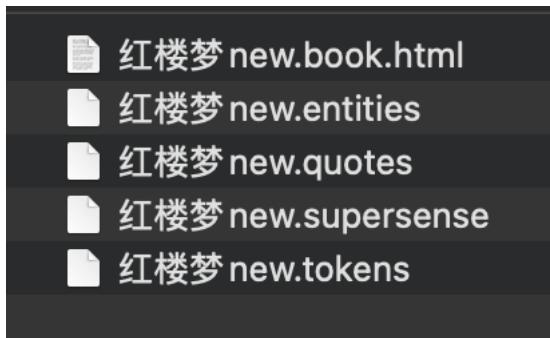


Figure 2

Among those outputs, quotes text and HTML are the two dominant files used in the project. In the former file, every sentence within the quotation marks in the book is extracted and assigned to a unique character ID, labeling the start and end positions to indicate the number of words (Figure 3). Meanwhile, the HTML file is more like a compound, including the lists of all named characters, organizations, and so on recognized by BookNLP(Figure 4), highlighting them clearly in the original text(Figure 5).

红楼梦 new.quotes						
quote_start	quote_end	mention_start	mention_end	mention_phrase	char_id	quote
72	76	103	103	he	1527	" spiritual understanding . "
106	192	103	103	he	1527	" In this busy , dusty world , having accomplished nothing , I suddenly recalled all the girls I had known , considering each in turn , and it dawned on me that all of them surpassed me in behaviour and understanding ; that I , shameful to say , for all my masculine dignity , fell short of the gentle sex . But since this could never be remedied , it was no use regretting it . There was really nothing to be done . "
292	407	409	409	you	1534	" Though my home is now a thatched cottage with matting windows , earthen stove and rope - bed , this shall not stop me from laying bare my heart . Indeed , the morning breeze , the dew of night , the willows by my steps and the flowers in my courtyard inspire me to wield my brush . Though I have little learning or literary talent , what does it matter if I tell a tale in rustic language to leave a record of all those lovely girls . This should divert readers too and help distract them from their cares . That is why I use the other name Jia Yucun . "
650	728	631	632	The monk	1543	" You look like a precious object , but you still lack real value . I must engrave some characters on you so that people can see at a glance that you 're something special . Then we can take you to some civilized and prosperous realm , to a cultured family of official status , a place where flowers and willows flourish , the home of pleasure and luxury where you can settle down in comfort . "
734	743	775	776	The monk	1543	" May I trouble you to enlighten me , "
747	768	775	776	The monk	1543	" as to what wonderful merits you will bestow on me ? And where do you mean to take me ? "
769	774	775	776	The monk	1543	" Do n't ask . "
779	789	775	776	The monk	1543	" You 'll find out all in good time . "
1034	1153	1026	1028	The Reverend Void	416	" Brother Stone , you seem to think that your tale recorded here is interesting enough to merit publication . In my view , in the first place , there is no way of finding out the dynasty and the year ; in the second , there is nothing here about worthy and loyal ministers and how they regulated the government and public morality . There are merely some girls remarkable only for their passion or folly , or

Figure 3

Named characters

10510 [NARRATOR]
7003 Baoyu (217)/Daiyu (976)/Aunt Xue (421)Presently Baoyu (3)/Soon Baoyu (2)/cousin Daiyu (2)/Only Baoyu (2)/In dismay Baoyu (1)/Fortunately Baoyu (1)/Only Daiyu (1)/Aunt Xue Conforts Daiyu Baoyu (1)/Aunt Xue who soothed her and finally persuaded her to stop weeping and go over to the other mansion (1)/Nowadays Baoyu (1)/So Daiyu (1)/Smilingly 4848 Xifeng (107)/Lady Wang (815)/Caiyun (47)/Jinchuan (34)/Wang Xifeng (5)/Only Xifeng (3)/Wang (3)/And Xifeng (2)/Presently Xifeng (1)/Lady Wang's maids Jinchuan and Caiyun who because they wanted to go were attending on Xifeng today (1)/eyen Lady Wang (1)/2741 Xifeng (1)/This lady Wang (1)/Lady Wang h- ad (1)/So Xifeng (1)/beyong Xifeng (1)/This min Xifeng (1)/Poor Xifeng (1)
7203 Baochai (959)/Master Bao (99)/Miss Baochai (56)/Cousin Baochai (38)/Bao (4)/Presently Baochai (2)/Cousin (2)/Even Baochai (1)/So Baochai (1)/cousin Baochai (1)/our Master Bao to Miss Baechai (1)
2141 Xieren (485)/Sister Xiren (15)/Miss Xiren (5)/Sister (4)/Little Sister (1)/Presently Xiren (1)/Pachyna Cooco Xires (1)/And Xire (1)/poor Xiren (1)/sister Xiren (1)/xiren (1)
2172 Jia Zheng (631)/Jia Shun (38)/Jia Shun (26)/Jingling (4)/Prefect Fu (3)/Marquis Shu (2)/Marquis Shu of Jingling , by whom he was sent, has she and Jia Zheng (1)the third son of Jia Shun (1)/young Marquis Shu (1)/Zheng (1)/Administration Jia Zheng (1)/The Prefect of Ping'an Jia Zheng (1)/Shu (1)
1509 Jia Lian (540)/Lian (23)/Lian (13)/Cousin Lian (1)/Lian (1)an (2)/Cousin Jia Lian (1)/Jia Lian outside (1)/Unhappy Jia Lian (1)/Jia Lian who had never liked him (1)
1177 Pinger (590)/Miss Pinger (17)/Sister Pinger (1)/Pinger Play Ting Pingyue (1)/Witching Pinger (1)/Tactful Pinger (1)/Only Pinger (1)/Presently Pinger (1)/Here Pinger (1)/Lucky Pinger (1)
1178 Baoyu (217)/Zhen Yuan (28)/Cousin Baoyu (4)/Zhen (4)/Uncle Baoyu (2)/young Zhen (2)/that Zhen Baoyu who they 's always talking about (1)/Baoyu (1)/this Chen Baoyu (1)
1085/1086 xing (239)/Xing (232)/and Xiting (1)/this hui Xifeng (1)/Lady Xing Feeling Weptong Pu Xifeng (1)/Xing (1)/Presently Xifeng (1)/Xifeng (1)
837 Xiangyun (333)/Cui (29)/Xiangyun (99)/Qianxue (1)/hui Xifeng (4)/Cui who went with Miss Xiangyun, Keren and Linchuan who 've died, Qianzhe who 's left, and the two of us
757 Xiangyun (1)/Presently Xiangyun (1)/Now Xiangyun (1)
757 Qingwen (295)/Sister Qingwen (19)/Only Qingwen (1)/ming Qingwen (1)/Now Qingwen (1)
754 Granma (170)/Liu (100)/Cutting Grannys Liu short Xifeng (1)/granny (1)/Liu (1)/Granny (1)
755 Xue Pan (151)/Xue Pan (51)/Liu Xiangtan (17)/Xue Pan (76) (6)/a certain Liuxian, whom Xue Pan had been longing to meet again ever since he first made his acquaintance
(1)/little Xue Pan (1)
610 Jizhan (277)/Sister Zijuan (10)/Miss Zijuan (3)/Sister (1)/So Zijuan (1)
605 Tanchun (385)/Mansion (25)/Presently Tanchun (1)
603 Yuanyang (319)/Sister Yuanyang (1)/Still Yuanyang (1)/Presently Yuanyang (1)/Sister (1)/Luckily Yuanyang (1)
544 Yuyun (108)/Jia Yucun (22)/Mr. Jia Yucun (2)/Mr. Jia Yucun (1)/This Mr. Jia Yucun (1)/Jia Yucun (1)/Jia Yucun!?" (1)
543 Yuanyang (139)/The Lady Dowager (67)/Lady Dowager (21)/Lady Dowager (1)/The Lady Dowager , who hastily stopped them and sent them off to have some refreshments in another room (1)
540 Jia Yun (135)/Master Yun (11)/Yun (10)/ha Yun (6)/she Jia (1)/young Master Yun (1)/Mr. Jia Yun (1)
486 Jia Rong (170)/Jia (43)/Rong (20)/ha Rong (9)/she Jia (4)/the Rong (1)/biao Rong (1)/Jia bing (1)/Jia zheng (1)
487 Jia Chen (232)/Zhen (14)/Cousin (7)/Zhen (14)/she Jia (8)/Ning (3)/the Ning (2)/Jia Chen of the Ning Mansion (1)/Jia She (1)/biai Zhen (1)/cousin Jia Chen (1)/Jia Mission (1)
482 Jia Rong (94)/Haun (15)/Jia Haun (15)/Presently Jia (1)
484 Xiangling (161)/Sister Xiangling (1)/Jia radish Xiangling (1)/Only radish Xiangling (1)/Xiang Ling (1)/Lovely Xiangling (1)
439 Xichun (200)/Miss Xichun (14)/young Xichun (1)/iuchun (1)/Flushing crimson Xichun (1)
426 Madam Lian (93)/Madam (91)/lina (1)/lina (10)/Madam (5)/Madam (1)/lina (1)/Good Madam Lian (1)/cousin Lian (1)
409 Li Wan (289)/Sun (16)/Biye (15)/Li Wan's maids Sun and Biye (1)/Li wan (1)/Li (1)
348 Yingchen (171)/Miss Yingchen (6)/Cousin Yingchen (4)/Soon Yingchen (1)/poor Yingchen (1)/Take Yingchen (1)

Figure 4

Text

Chapter 1

[Zhen Shiyin]412-shiyin in a Dream Sees the Jade of Spiritual Understanding [Jia Yucun]413-yucun in [His]413-yucun Obscurity Is Charmed by [a Maid]1525-a main

This is the opening chapter of the novel. In writing this story of the Stone [1526-the stone] the author [1527-the author] wanted to record certain of his [1527-the author] past dreams and illusions , but he [1527-the author] tried to hide the true facts of his [1527-the author] experience by using the allegory of the jade of "Spiritual Understanding". [1527-the author] Hence his [1527-the author] recourse to names like Chen Shiyin.⁴ [144-shen why] But what are the events recorded in this book , and who are the characters [1528-the characters] ? About this [1527-the author] said : "In this busy , dusty world [2738]-this busy , dusty world , having accomplished nothing , [I] [1527-the author] suddenly recalled all the girls [I] [1527-the author] had known [2738]-all the girls i had known , considering each in turn , and it dawned on me [me] [1527-the author] that all of them [2738]-all the girls i had known surpassed [me] [1527-the author] in behaviour and understanding ; that [I] [1527-the author] shameful to say , for all my [1527-the author] masculine dignity , fell short of the gender sex . But since this could never be remedied , it was no use regretting it . There was really nothing to be done . [1527-the author] I [0-4] decided then to make known to all how [I] [0-4] , though dressed in silk and delicately nurtured thanks to the Imperial favour and my [0-4] ancestors [1529-my ancestor] virtue , had nevertheless ignored the kindly guidance of my [0-4] elders [1530-my elders] as well as the good advice of teachers [1531-teachers and friends] [1532-friends] , with the result that I [0-4] had wasted half my [0-4] life and not acquired a single skill . But no matter how unforgivable my [0-4] crimes , I [0-4] must let not all the lovely girls I [0-4] have known [1531-all the lovely girls i have known] pass into oblivion through my [0-4] wickedness or my [0-4] desire to hide my [0-4] shortcomings . "Though [my] [1534-my home] [2738]-my home is now a thatched cottage with matting windows , earthen stove and rope -bed [2739] a thatched cottage with matting windows , earthen stove and rope -bed , this shall not stop me [1534-my home] from laying bare [my] [1534-my heart . Indeed , the morning breeze , the dew of night , the willows by [my] [1534-my steps] and the flowers in [my] [1534-my courtyard] [2739]-my courtyard inspire [me] [1534-my to wield [my] [1534-my brush . Though [I] [1534-my] have little learning or literary talent , what does it matter if [I] [1534-my] tell a tale in rustic language to leave a record of [all those lovely girls] [2739]-all those lovely girls ? This should divert [readers] [2739]-readers too and help distract [them] [2739]-readers from their [2739]-readers cares . That is why [I] [1534-my] use the other name [Jia Yu'en] [413-yeung] -[2734-my] Do you [1534-my] know , "Worthy Readers [1535-worthy readers] , where this book comes from ? The answer may sound fantastic , yet carefully considered is of great interest . Let me [0-4] explain , so there will be no doubt left in your [1534-my] minds . When the goddess Nu Wa [1536-the goddess nu wa] melted down rocks to repair the sky , at Baseless Cliff [1-baseless cliff] in the Great Waste Mountain [1537-the great waste mountain] she [1536-the goddess nu wa] made thirty - six thousand five hundred and one blocks of stone , each a hundred and twenty feet high and two hundred and forty feet square . She [1536-the goddess nu wa] used only thirty - six thousand five hundred of these and threw the remaining block down at the foot of Blue Ridge3 Peak [2-blue ridge3 peak] [1538-the foot of blue ridge3 peak] . Strange to relate , this block of stone after tempering had acquired spiritual understanding . Because all

Figure 5

The mechanism of applying BookNLP's analysis is to retrieve Lin Daiyu's

character ID, which was assigned by BookNLP automatically, and then extract every sentence aligning with the ID. Thus, the text output with the content of Lin Daiyu's discourse, the first dataset, should be gained.

As to the second dataset, compared with the first one, it is easier to manipulate.

To begin with, a Python code using the re (regular expression) module was written to extract discourses from the book, and then the extracted sentences were saved into a

new text file. After that, Python is applied to eliminate Lin Daiyu's part from the overall output gained in the first step. Therefore, the second dataset was achieved successfully.

Voyant

Word Frequency

This project will first conduct a comparative analysis of the discourse of Lin Daiyu and that of the rest of the characters from the perspective of word frequency. The purpose is to analyze the differences in the keywords used by Lin Daiyu and other characters in discourse, which represents their wording preferences and emotional tendencies in the novel.

The Discourses of Lin Daiyu

Firstly, we created a word cloud with Voyant's 'Cirrus tool' and generated Figure 1a. Moreover, the Stopwords function is used to remove common words that have little practical meaning, such as 'it's,' 'I'm,' 'I'll,' and 'I've,' to make keywords in the text more prominent.

From Figure 1a, we can see the frequency of word occurrence. The larger the word, the higher the frequency of word occurrence. In addition, the keywords of the text are shown clearly: 'don't,' 'come,' 'want,' 'cousin,' 'master,' etc. By observing these keywords, we can perceive some themes or topics. For example, 'cousin' and 'master' express family relationships and class gaps respectively, 'don't' has an apparent negative meaning, and verbs such as 'want' and 'come' have action

meanings.



<https://voyant-tools.org/?corpus=870a41d2b6ffacec762f7da7d450ccb1&stopList=keywords-3adf91cb87e39bb85fb2ee60becb635b&visible=500&view=Cirrus>

Figure 1a

Then, Voyant's 'Terms tool' is applied to create a table of the top 15 most frequently used terms (Figure 1b).

In Figure 1b, the 15 most frequently used words and their frequency of occurrence are exhibited transparently.

		Term	Count	Trend
⊕	□	1 don't	22	
⊕	□	2 come	20	
⊕	□	3 want	14	
⊕	□	4 cousin	13	
⊕	□	5 like	12	
⊕	□	6 time	11	
⊕	□	7 better	11	
⊕	□	8 right	10	
⊕	□	9 tell	9	
⊕	□	10 n't	9	
⊕	□	11 baochai	9	
⊕	□	12 bao	9	
⊕	□	13 means	8	
⊕	□	14 master	8	
⊕	□	15 ll	8	
⊕	□	16 leave	8	

<https://voyant-tools.org/?corpus=870a41d2b6ffacec762f7da7d450cbb1&palette=23bd039e8831797f3c885db4b44460d9&stopList=keywords-3adf91cb87e39bb85fb2ee60becb635b&comparisonCorpus=&view=CorpusTerms>

Figure 1b

In Figure 1b, the 15 most frequently used words and their frequency of occurrence are exhibited transparently.

To begin with, an interesting discovery attracted our attention. The most frequent word is ‘don’t’ (22 times), which has a strong negative meaning, implying Lin Daiyu’s rejection or loathness of certain things. We cannot help but think about this question: Under what circumstances does she have such emotions? Within the scope of this corpus (Lin Daiyu’s discourses), the people she talks to are other characters within the novel and herself. While talking to others, ‘don’t’ means that she does not agree with the other person’s point of view or asks the other person not to do something; while talking to herself, ‘don’t’ implies Lin Daiyu’s conflicting psychology and complex emotions. When a character ruminates on questions like ‘do’ and ‘don’t,’ her feelings are deep and complex. In short, ‘don’t’ reflects Lin Daiyu’s rebellious spirit of unwillingness to be bound or controlled by others, and it also shows her attitude of expressing her independent thinking.

In addition, ‘want’(14 times) and ‘like’(12 times) also appear frequently. This group of words has an obvious emotional tendency, indicating that Daiyu directly expresses her thoughts for likes and dislikes. Moreover, ‘cousin’(13 times) and ‘master’(8 times) are appellations indicating family and social status. In *THE DREAM OF THE RED CHAMBER*, most of the people close to Lin Daiyu in age are her

cousins, such as ‘baochai’(9 times) and ‘bao’(9 times), who are the two other main characters in this novel. This phenomenon also shows that they have active contacts and close communications. Within the list in Figure 1b, ‘come’ (20 times) and ‘tell’ (9 times) are verbs occurring in conversations, implying that this is a large family with close family ties. Meanwhile, it is not hard to conjecture that this kind of big family lifestyle was relatively common in the era when the author created this work, providing vital information on the social background of the work. Besides that, ‘master’ is used as an honorific appellation for men with great power in this book and is a word with strong class attributes. From the analysis in the previous, we know that Daiyu is a woman with independent thoughts. We found that the word ‘aunt’ conquers a very prominent position in the word cloud in Figure 1a, indicating that for the female elders in the family, Lin Daiyu chose to use ‘aunt’, which expresses kinship relationships, instead of names referring to class gaps. This shows that in Lin Daiyu's heart, she regards female elders as family members rather than leaders. Nevertheless, although she possesses rebellious consciousness and strives to struggle with the toxic class-clear social common traits, she still has to abide by the etiquette of the feudal era and use honorifics in her daily conversations, making her miserable portrayal even deeper.

The Discourses of The Rest

In this part, the same function of Voyant was used to analyze the discourses of other characters in *THE DREAM OF THE RED CHAMBER*, generating Figure 2a and Figure 2b. Based on the visual results, we will conduct a comparative analysis.



<https://voyant-tools.org/?corpus=210f19ce5e31f5fa421712238656b35b&stopList=keywords-fb6b8b10bda75351d2bc9598558bf454&visible=500&view=Cirrus>

Figure 2a

		Term	Count	Trend
+	<input type="checkbox"/>	1 just	301	
+	<input type="checkbox"/>	2 good	295	
+	<input type="checkbox"/>	3 come	248	
+	<input type="checkbox"/>	4 old	247	
+	<input type="checkbox"/>	5 like	238	
+	<input type="checkbox"/>	6 master	237	
+	<input type="checkbox"/>	7 madam	224	
+	<input type="checkbox"/>	8 know	222	
+	<input type="checkbox"/>	9 time	187	
+	<input type="checkbox"/>	10 make	177	
+	<input type="checkbox"/>	11 tell	170	
+	<input type="checkbox"/>	12 let	163	
+	<input type="checkbox"/>	13 lady	163	
+	<input type="checkbox"/>	14 right	162	
+	<input type="checkbox"/>	15 better	158	
+	<input type="checkbox"/>	16 things	153	

<https://voyant-tools.org/?corpus=210f19ce5e31f5fa421712238656b35b&stopList=keywords-fb6b8b10bda75351d2bc9598558bf454&comparisonCorpus=&view=CorpusTerms>

Figure 2b

We still start with analyzing higher-frequency words. ‘just’ appears 301 times in total. This is a word with various meanings, demanding for the context to do accurate

analysis. Thus, this word will be analyzed later.

Above all, there is a clear difference in the discourses of other characters from Daiyu's: the words representing class were ranked higher and more numerous. First, the ranking of 'master' (237 times) becomes higher. In Lin Daiyu's discourses, 'master' ranks 14th, whereas it rises to sixth in the other's. This increase indicates that other characters in *THE DREAM OF THE RED CHAMBER* are more class-conscious compared to Lin Daiyu, meanwhile showing that Lin Daiyu has a closer connection with people sharing the same class, implying that she does not want to have more contact with those in power. Furthermore, 'madam' (224 times) and 'lady' (163 times) are newly added high-frequency words. These two are honorific names for female people in power or female elders. The emphasis on the appellations of people from different social or family levels suggests that in the entire novel of *THE DREAM OF THE RED CHAMBER*, the concepts of superiority and class are comprehensive and substantial, and both male and female elders have specific titles. However, these two class-clear words hardly appear in Daiyu's frequent usage words, confirming the former statement that Daiyu wishes to revolt and she truly does her best to achieve her wish, making her stand out from other characters within the book.

In addition to these differences, the discourses of the other characters also have some similarities with Lin Daiyu's. For example, verbs with communicative meanings, such as 'come'(248 times), 'tell'(170 times), 'let'(163 times), and 'make'(177 times), also appear frequently. This supports the above analysis from the perspective of other characters: this is a large family with closely connected family

members, reflecting the social background of the novel.

Word Collocation

In the above content, we have already illustrated a brief understanding by combing the word frequency in the corpus and achieved a preliminary exploration of Daiyu's emotional expression. Next, we will analyze the collocations of these keywords based on some of the above keywords and further analyze the corpus. This section aims to examine the frequency of keywords co-occurring with other words and their positions in the corpus, which will also allow us to study Daiyu's language style and characteristics from another aspect further. Meanwhile, the same function of Voyant is applied to the corpus of other character discourses, using a comparative method and analyzing it in conjunction with our research questions.

The Discourses of Lin Daiyu

As shown in Daiyu's word cirrus in the previous section, among the most frequently occurring words, "don't" is an obviously negative word; "want" and "like" are two words that explicitly express emotions. These words are not only keywords to express emotions, but also how they are matched may reveal the character's mental state in a specific situation. For example, the collocation of the word "don't" may reveal the character's resistance to certain things or behaviors or imply inner psychological conflicts. Additionally, interactions between characters and other individuals may be involved as well. The word combination of "want" and "like" may express the characters' preferences, desires, or the interaction between characters. To

begin with, we will concentrate on the collocations of the word "don't," compiling a list of the most frequent collocations and analyzing the potential messages that these collocations may convey.

Term	Collocate	Count (context) ↓
don't	want	10
don't	i'm	8
don't	don't	8
don't	come	8
don't	like	7
don't	tell	6
don't	say	6

268 context

<https://voyant-tools.org/?context=30&query=don%E2%80%99t&corpus=59bb4f9922393f78e5b32b7ba7bb3d49&view=CorpusCollocates>

Figure 1a

As shown in Figure 1a, the most frequent collocations with "don't" in the Daiyu Corpus are "want, I'm, don't, come, like." The collocation of "don't want" commonly indicates preferences and desires. Considering this result in conjunction with our corpus, we have arrived at two potential interpretations. Firstly, "don't" directly connects with "want." Through this expression, we can infer that Daiyu is a character who straightforwardly expresses her wishes and preferences in specific contexts. Further interpretation can be derived by combining this with the description of her background in the original work. Daiyu comes from a scholarly family but lost her mother at a young age and was brought to the Jia family by her maternal grandmother. In the third chapter, Daiyu is described as follows: 'Although Lin Daiyu, who is 'carefree and willful,' first came to Jia's house, she always remembered her mother's

teachings. Be careful every step of the way, be careful at all times, don't say one more word, don't take one more step, for fear of being laughed at.'(Cao,2014). However, Daiyu did not truly put her mother's teachings into practice. She continued to speak her mind. With her intelligence, she must have been aware of the intricate relationships within the Jia family, but she simply disdained getting involved (Zhu,1994). From this, it becomes apparent that the author has portrayed Daiyu as a character with a pure temperament. Secondly, another situation is that these two words are not directly collocated but appear separately in different sentences. Their proximity in the corpus does not provide significant information in terms of their relationship.

" When you see the old lady , please give her my respects and tell her I 'm a bit poorly , but it 's nothing serious she 's not to worry . "

" Whom were you talking to ? "

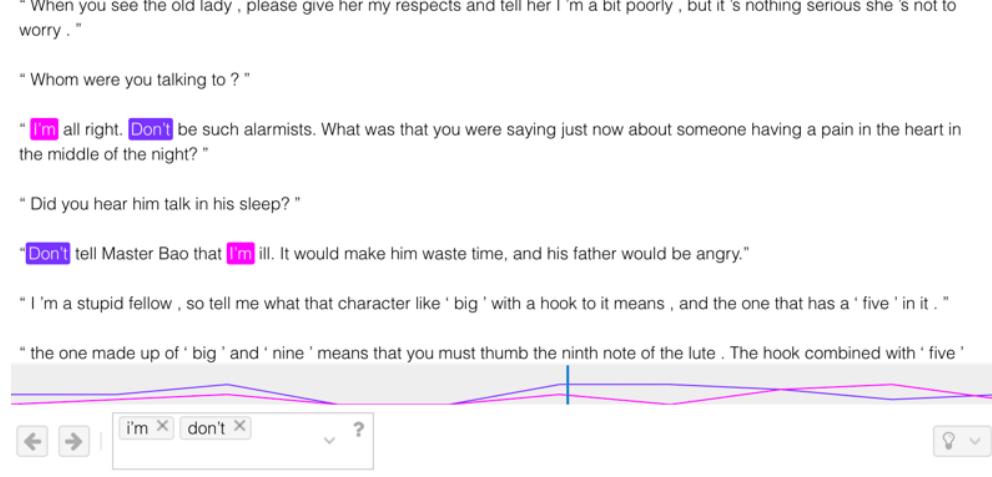
" I'm all right. Don't be such alarmists. What was that you were saying just now about someone having a pain in the heart in the middle of the night? "

" Did you hear him talk in his sleep? "

"Don't tell Master Bao that I'm ill. It would make him waste time, and his father would be angry."

" I'm a stupid fellow , so tell me what that character like ' big ' with a hook to it means , and the one that has a ' five ' in it . "

" the one made up of ' big ' and ' nine ' means that you must thumb the ninth note of the lute . The hook combined with ' five '

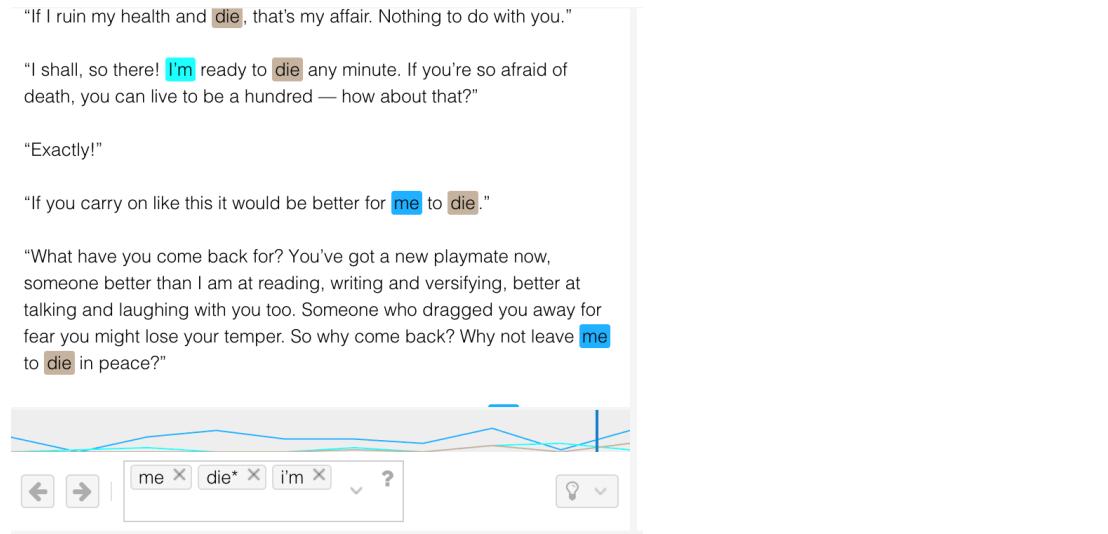


<https://voyant-tools.org/?start=999&skipToDocId=4abaadb3e803f30833c8cc5b1e5b8f54&corpus=59bb4f9922393f78e5b32b7ba7bb3d49&view=Reader>

Figure 1b

The collocations "don't" and "don't," as well as "I'm" and "don't," are not common lexical combinations(Figure 1a). In order to explain this phenomenon, we combine Voyant's instructions for the word collocation function. Corpus collocations

are a tabular view of which terms occur more frequently near keywords in the entire corpus. In other words, near don't, the two words I'm and don't appear frequently. Nevertheless, after utilizing the reader function in Voyant and analyzing our corpus, we found that "I'm" and "don't" are merely positioned closely to each other in the text and cannot be considered as one of the factors for analysis (Figure 1b). The same situation applies to the collocation between "don't" and "come." As for the collocation between "don't" and "like," we found that such a combination only appeared in one sentence. Considering that a single example is insufficient to support an inference about the character's personality, we refrain from making an interpretation in this regard.



<https://voyant-tools.org/?start=1998&skipToDocId=4abaadb3e803f30833c8cc5b1e5b8f54&corpus=59bb4f9922393f78e5b32b7ba7bb3d49&view=Reader>

Figure 1c

When searching with "want" as the keyword, there is one word that caught our attention, which is "die" (Figure 1c). "Die" appears a total of six times in the corpus, with two occurrences of "want" and "die" appearing together and five occurrences of

"die" appearing with "I" or "me." Such collocations directly reflect Daiyu's negative emotions. Imagine, when the author depicts the character's discourse and directly uses the word "die," what kind of character information does he intend to convey?

Considering the context of the novel, we speculate that when Daiyu utters such sentences, she does not actually desire death but rather adopts a "making a fuss about nothing" approach to affirm her position in Baoyu's heart. It can be seen as a way for the two of them to engage in playful banter. However, this pessimistic style of flirting is unique to Daiyu. From these collocations, we can perceive that the Daiyu portrayed by the author is a pessimist.

The Discourses of The Rest

When analyzing the word collocations of Daiyu, we selected the keywords "want" and "don't" to gain a deeper understanding of her character traits. To further highlight the uniqueness of the character, we decided to apply these two keywords to the corpus of other characters as well, focusing on the differences in collocations between the two corpora.

Term	Collocate	Count (context)
don**	right	59
don**	worry	58
don**	i've	56
don**	say	53
don**	tell	52
don**	time	50

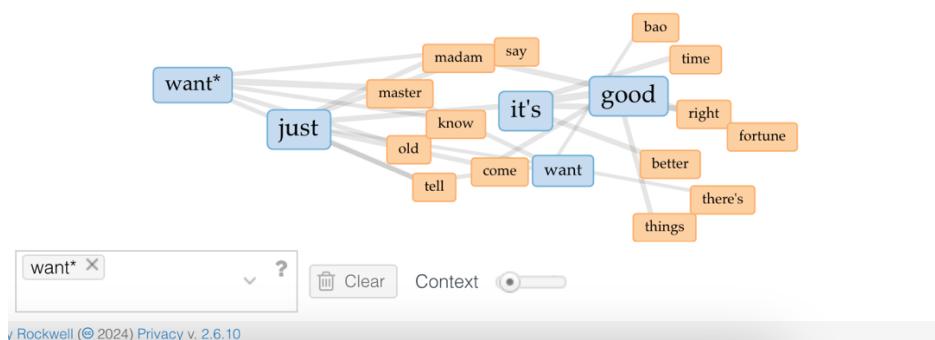
don't don** ? 2,437 context

https://voyant-tools.org/?corpus=8f9658106270867dd80f5ae4c20de31e&context=30&query=don%27t&query=don%27*&view=CorpusCollocates

Figure 2a

While examining the collocations related to "don't," we noticed one collocation that did not appear in the Daiyu corpus: "don't" and "worry." (figure 2a). This phrase, meaning "do not worry," appeared thirty times throughout the entire corpus. "Don't worry" is a phrase used to express comfort and reassurance, and its number of occurrences in the corpus may indicate that other characters use this phrase more often, which could reflect their attitudes when facing emotions or pressures.

Considering the frequent appearance of this phrase and comparing it with the Daiyu corpus, we have two interpretations: Firstly, regarding emotional expression, Daiyu may be more inclined to express her feelings and use comforting phrases less often. This aspect reflects her candid nature. Secondly, Daiyu tends to suppress her emotions. While other characters use such phrases to comfort others, Daiyu's limited usage shows that she is less inclined to display her concerns openly and prefers to deal with emotional burdens alone.



https://voyant-tools.org/?query=want*&mode=corpus&corpus=8f9658106270867dd80f5ae4c20de31e&view=CollocatesGraph

Figure 2b

In the corpus of discourse from other characters, the word "want" appears in collocations with "master" and "madam" (figure 2b), which do not occur in Daiyu's

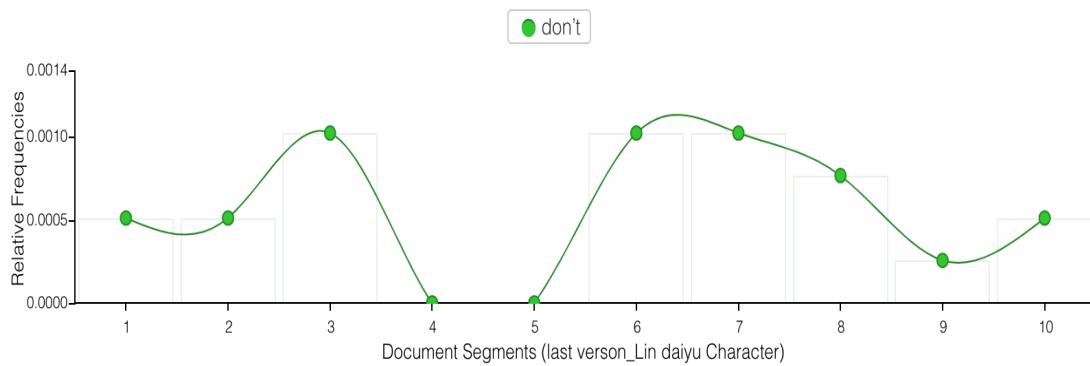
corpus. After analyzing the corpus, we interpret the occurrence of these results as follows. Firstly, in the first case where "want" appears together with "master" or "madam" in a sentence, such sentences usually occur when inquiring about what the master or madam desires. The second situation is where the three words appear in separate sentences but in proximity within the corpus. Considering both situations and comparing them with the analysis results from Daiyu's corpus, we consider this to reflect the adherence to power hierarchies among the members of the Jia family.

Although "master" appears as one of the frequent keywords in Daiyu's corpus only eight times, and "madam" appears only twice, we infer that Daiyu is not inclined to display obedience to authority. She may be more focused on her emotional inner world than social status or service relationships, as her corpus contains more vocabulary directly expressing emotions.

Deeper Analysis Combined with Voyant

After getting the analysis results of the word frequency and word collocation of the character discourses in THE DREAM OF THE RED CHAMBER, we will use Voyant's "Trends Tool" and "Contexts Tool" to put them into the text, hoping to find answers to the research questions from a broader perspective.

The Distribution Trend of 'Don't' in The Whole Book and The Changes in Lin Daiyu's Rebellious Spirit



<https://voyant-tools.org/?corpus=91660373d8751b6a344e752734837a85&stopList=keywords-94c6e7a1df396049d8635c422fd1cea8&query=don%E2%80%99t&mode=document&view=Trends>

Figure 3a

First, we used the ‘Trends Tool’ to generate a graph of the frequency of ‘don’t’ occurrences throughout the document (Figure 3a). The x-axis in the figure means Document Segments, which divides Lin Daiyu’s discourses into ten equal parts, from 1 to 10, in line with the development order of the novel, and the y-axis means Relative Frequencies. Next, we will analyze the changing trends from this picture further.

Document	Left	Term	Right
last ver...	so boisterous and pert ? " " You	don't	know her yet . " " Thank you
last ver...	me see to it . " " I	don't	want them . They've been handled
last ver...	were an Imperial edict." "You	don't	understand, aunt," "It doesn't matter
last ver...	behave in this outrageous fashion." "	don't	be such a spoil-sport
last ver...	ever get from me." "You	don't	have to treat me like
last ver...	declared. "First you say you	don't	want it, and now you're
last ver...	really blush for you." "Why	don't	you go for a stroll
last ver...	gold to match it. So	don't	you have a warm scent
last ver...	you to bring me ? " " I	don't	want it now . Put it
last ver...	What 's the matter ? " " Cousin ... " "	don't	fret yourself . It 's only
last ver...	talking to ? " " I'm all right.	don't	be such alarmists. What was

<https://voyant-tools.org/?corpus=91660373d8751b6a344e752734837a85&stopList=keywords-94c6e7a1df396049d8635c422fd1cea8&query=don%E2%80%99t&view=Contexts>

Figure 3b

Secondly, we used the ‘Contexts Tool’ to get Figure 3b where ‘don’t’ appears in

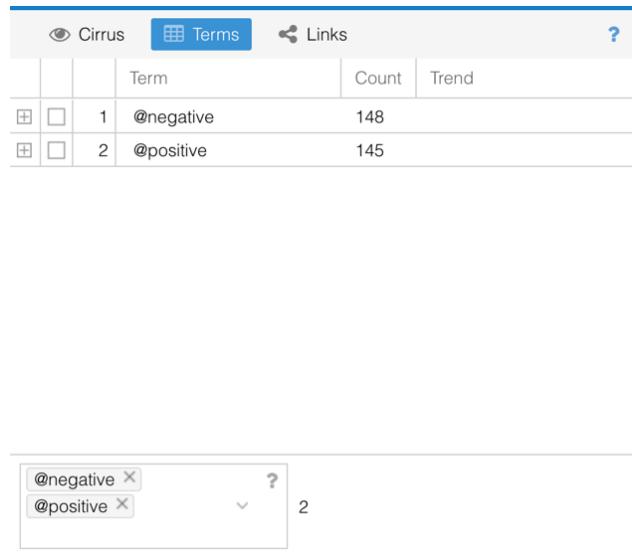
the original text.

We have already analyzed the rebellious spirit of Daiyu represented by ‘don’t’ in word frequency. Now look at how it’s changing. The values start low and vary widely overall. This shows that Daiyu did not have a strong spirit of rebellion from the beginning but gradually awakened after observing and experiencing life. In addition, this rebellion also shared a weakening and fluctuating stage, which illustrates Daiyu's contradictory and worrying character. The first peak was reached in Part 3. This part corresponds to Chapters 24-36 of the book. In this part, the Jia family, a feudal family, established a utopia called Grand View Garden (Cao,2014). In this utopia, all the cousins who are similar to Daiyu live without the interruption of the adult world. They meet every day, write poems, and engage in literary creation. It was this free environment that gradually inspired Daiyu's awareness of self-awareness and social identity, and she realized that the feudal education she passively accepted was not necessarily correct.

In addition, parts 6 and 7 are the second peak, corresponding to chapters 60-84 of the novel. In this part, the emotional line between Lin Daiyu and the male protagonist, Baoyu, gradually unfolds. The two of them can be said to be soul mates, but some misunderstandings about daily trivial matters existed between them. For example, we can see the contexts from Figure 3b: ‘You don’t understand,’ ‘I don’t know,’ and ‘I don’t want.’ Therefore, the girlish feeling of first love is revealed at the stage of generating and eliminating misunderstandings. The lover's incomprehension of herself is also the main stage of Daiyu's awakening of self-awareness.

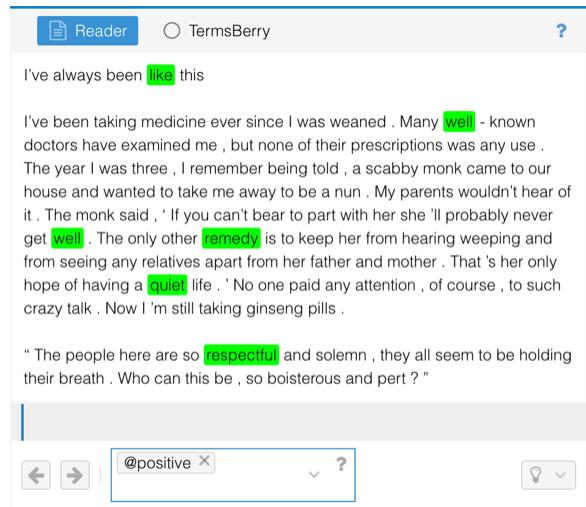
Decoding Lin Daiyu's Emotion by Examining the Words with Positive Connotations

Connotations



<https://voyant-tools.org/?query=%40negative&query=%40positive&corpus=59bb4f9922393f78e5b32b7ba7bb3d49&view=CorpusTerms>

Figure 3c



<https://voyant-tools.org/?skipToDocId=4abaadb3e803f30833c8cc5b1e5b8f54&corpus=59bb4f9922393f78e5b32b7ba7bb3d49&view=Reader>

Figure 3d

		Term	Count	Trend
1		don't	22	
2		come	20	
3		want	14	
4		i've	14	
5		i'm	13	
6		it's	13	
7		cousin	13	
8	<input checked="" type="checkbox"/>	like	12	
9		time	11	
10	<input checked="" type="checkbox"/>	better	11	
11	<input checked="" type="checkbox"/>	right	10	
12		tell	9	

<https://voyant-tools.org/?corpus=59bb4f9922393f78e5b32b7ba7bb3d49&view=CorpusTerms>

Figure 3e

To explore Daiyu's corpus further, we applied the "Categories" feature in Voyant, which allows us to label most of the words in the corpus based on their positive and negative connotations using the built-in category tags (@positive and @negative). By entering "@positive" in the search bar, the total amount of vocabulary in the corpus labeled as having a positive connotation is displayed. With the help of this method, we examined the number of vocabulary items with positive and negative connotations in the corpus (Figure 3c). The difference in quantity between the two categories was not significant. Therefore, we shifted our view to the specific vocabulary items marked with positive connotations. Through the 'reader' tool in Voyant, we discovered that vocabulary items marked with positive connotations are highlighted with a green background (Figure 3d). Based on this observation, we used the 'terms' tool to search which specific words were marked as having positive connotations (Figure 3e). This

chart exhibits the vocabulary items labeled with positive connotations and their frequency of occurrence.

Figure 3e shows that "like," "better," and "right" are the three most frequently occurring vocabulary items marked with positive connotations. However, individual words may have different connotations depending on the context, whether negative, positive, or other meanings. Therefore, Daiyu's discourse must be considered in determining whether these words are used exclusively to express positive emotions in specific contexts. Firstly, we utilized the 'Context' tool to quickly locate and examine the specific sentences where the word "like" appeared. We found that out of these 12 sentences, only one of them used "like" to express Daiyu's positive emotions, conveying the meaning of "desire" or "wish." The rest of the occurrences were used in the meaning of "similar" or "resembling." Therefore, under the above situation, this keyword does not directly assist us in analyzing Daiyu's character traits.

Next, we applied the same method to search for the occurrences of "better." We discovered two primary expressions. The first one is, in the structure "you'd better," it indicates a recommendation or suggestion to do something. In this case, Daiyu expresses her dissatisfaction with life twice, including one instance where she utters a desire to die. The second expression is used in the comparison. Under such expression, we noticed an interesting phenomenon where Daiyu uses the word "better" to convey a sense of jealousy and inferiority. For instance, she applies "better" to compare herself with Baochai in the secrecy of her own heart, which is one of her female cousins, when Baoyu overlooked her and found a companion who was

more knowledgeable in literature and writing than her in his point of view. This usage indicates Daiyu's attitude toward her relationship with Baoyu while avoiding directly expressing dissatisfaction. From this, we can sense the sensitivity and inner turmoil in Daiyu's characteristics. Lastly, we conducted a search for "right" and found that its most common collocation in the corpus is "all right" (occurring six times). In other sentences, "right" was used to convey the meaning of correctness or indicate direction. Regarding this phenomenon, we consider "all right" a common colloquial expression that doesn't provide significant assistance in delving deeper into the analysis of the character's image.

After analyzing the corpus using the category tool, we found that the positive and negative could not be directly identified from the counts. The previous analysis confirmed that in the use of the word "better," Daiyu expresses her negative thoughts by using this word rather than positive emotions in the sentences, and her sensitivity was evident from the way she expressed them.

LIMITATIONS OF TOOLS

BookNLP

While reviewing the analysis results of BookNLP, we shockingly found a significant problem with entity division, leading to the wrong character recognition and ID allocation. The machines mistakenly recognized the two main characters in this book, which are respectively the representatives of male characters and female characters, Jia Baoyu and Lin Daiyu, as one person. Reflecting on outputs, Baoyu and Daiyu shared one collective character ID (Figure 1), and as a result, some of their discourses were mixed up.

	TAGID	TAGNAME	NOM	PER	you young people
2943	41213	41213	PRON	PER	you
548	41220	41220	PROP	PER	Daiyu
548	41224	41224	PRON	PER	her
2484	41225	41225	NOM	FAC	her room
588	41227	41228	PROP	PER	Mrs. Zhou
548	41230	41230	PRON	PER	her
548	41232	41232	PROP	PER	Baoyu
2943	41244	41244	PRON	PER	him
588	41246	41247	PROP	PER	Mrs. Zhou
2943	41249	41249	PRON	PER	her
588	41254	41254	PRON	PER	she
615	41260	41261	PROP	PER	Madam Xue
0	41263	41263	PRON	PER	me
2943	41266	41266	PRON	PER	you
548	41279	41279	PROP	PER	Baoyu
0	41283	41283	PRON	PER	me
2943	41288	41288	PRON	PER	He
2910	41306	41307	NOM	FAC	the Palace
548	41309	41309	PROP	PER	Daiyu
548	41315	41315	PROP	PER	Baoyu
0	41321	41321	PRON	PER	I
548	41329	41329	PRON	PER	she
2160	41335	41337	NOM	PER	the other girls
2957	41345	41349	NOM	PER	Each of the young ladies
2946	41347	41349	NOM	PER	the young ladies
0	41357	41357	PRON	PER	you
2203	41359	41359	NOM	PER	miss

Figure 1

In order to fix this unanticipated problem, the output of BookNLP must be processed by Python. Therefore, except for retrieving Daiyu's character ID, the

character name corresponding to the sentence was extracted to check if it equals Daiyu's name. With the participation of this new Python code, the overall output eventually was the extraction of Daiyu's discourses.

Having experienced the complete process of operating BookNLP and managed to work out the difficulties that occurred within, we discovered three overall problems with this tool. The first one, which also offers the most significant influence on this project, is the inaccuracy of character division. This wrong category directly affects the character ID allocation and quote matching. Because of that, users fail to extract specific characters' sentences only according to their ID, whereas they have to input more restrictions to help process the output text files again. Then the second problem is the inaccurate judgment of sentences belonging to the personal pronoun, for example, she, his, I, etc. Every time BookNLP identifies the occurrence of a personal pronoun, it will automatically allocate a character ID to it, given after its internal calculation to decide which character this sentence belongs to. This substantially ought to be a perfect assistant for literary analysis, making it more comprehensive and profound. Nevertheless, the wrong sentence attribution leads to narrowing down the available data to study character. Researchers are not able to use the data with the retrieved discourses by ID only but by adding the specific name to ensure the accuracy of output, which, in fact, results in the reduction of data compared to the former. Furthermore, this decline in available data for a study may cause qualitative or quantitative effects on the result. Lastly, the third problem is inaccurate judgment of a sentence's start or end position. For example, as shown in Figure 2, 'said Vermilion

Pearl' is recognized by BookNLP as within the discourse said by the character Vermilion Pearl, which turns out to be a narration the author wrote to connect two separate discourses. This kind of mistake occurs a lot in the final output and can hardly be manually found one by one for its random scattering. Therefore, it is inevitable that it affects the analysis of character discourse more or less.

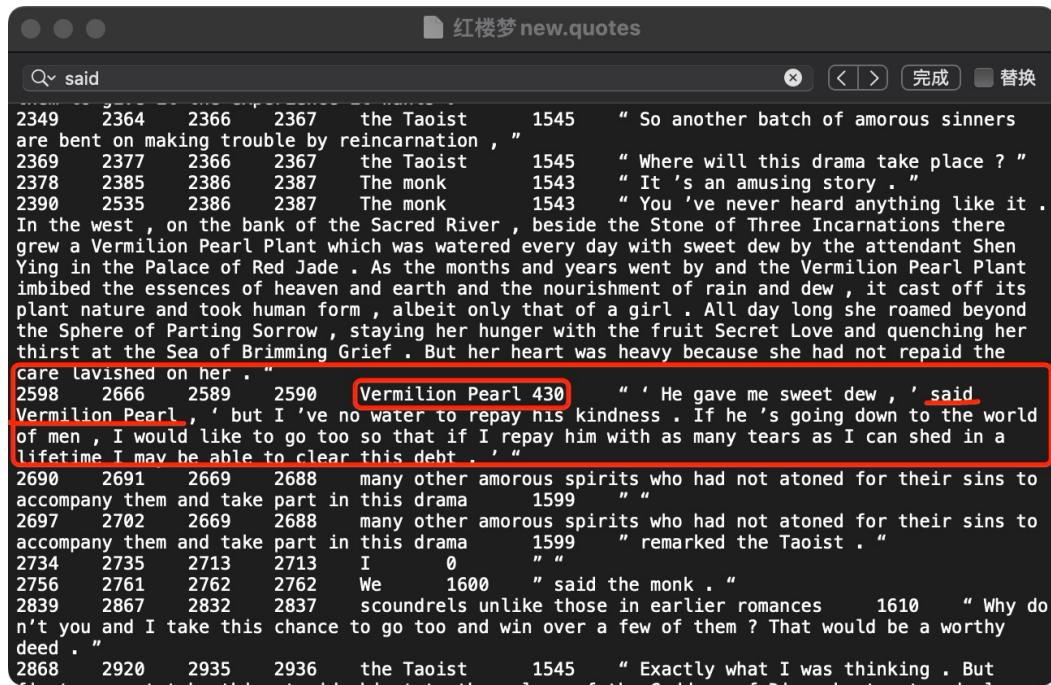


Figure 2

Overall, based on the previous demonstration, the dominant limitation of BookNLP is its lack of familiarity with different patterns of literary works as a book processing and analysis machine. Its module and calculation learning materials need to involve enough ranges of literary kinds to support doing accurate case analysis and then manage to improve those three problems mentioned before.

Voyant

By utilizing the functionalities within Voyant, we can quickly focus on the direction of analysis. Functions like word frequency and word collocations, which are provided based on the function of the tool, facilitate a rapid understanding of the corpus for researchers. Although our project leverages the text analysis tool Voyant to delve into the differences between Lin Daiyu's discourses and those of other characters, it is essential to acknowledge its limitations.

Firstly, Voyant is a distant reading tool capable of providing insights into word frequency, embeddings, and other statistical information. Still, it may not capture subtle contextual and tonal variations in language that are crucial for a nuanced language analysis. Given that our research aims to gain a deep understanding of Lin Daiyu's unique expression in discourses, these subtle aspects are integral to authentic language analysis.

Secondly, Voyant's text analysis primarily focuses on quantifiable and statistical data, exhibiting limitations in the nuanced understanding of context and the specific contextualization of emotions when analyzing emotional tendencies. This poses a challenge when interpreting Lin Daiyu's emotional expressions. The tool utilizes an embedded corpus to label vocabulary's positive or negative connotations. This may result in insufficient accuracy in the labeling. Although the tool has a functionality that allows us to define the nature of these vocabulary items ourselves, considering the polysemous nature of many words in English, this is also a limitation of the tool.

Moreover, while Voyant can offer contextual information for key terms, its understanding of complex discourse structures and grammatical features might need to be improved. For our project, this limitation may result in an incomplete understanding of certain aspects of Lin Daiyu's language style and emotional tendencies.

Therefore, while Voyant provides valuable overall data for our research, it is crucial to approach its limitations cautiously. Integrating other in-depth language analysis methods to understand Lin Daiyu's discourse characteristics may be beneficial comprehensively.

CONCLUSION

Through the analysis of word frequency and word collocation, Lin Daiyu's character profile has taken shape. Under the most frequently used words, Daiyu's inclination to exhibit unwillingness and dislike more than positive expressions; meanwhile, her avoidance of using class-clear appellations has been revealed as well. In terms of these two discoveries, we preliminarily draw a conclusion that Lin Daiyu holds a rebellious spirit, which is a life value out of step with contemporary society. Then, from the analysis of word collation, Daiyu's personality was explored further. The straightforward and pure temper inside builds a unique and courageous woman image that makes her stand out not only in the book but also in the era. Nevertheless, the high-frequency negative expression of Daiyu attracted our attention as well. She expresses the wish to die more than once, which induces our curiosity to do more profound study. Carrying two main thoughts, we advanced into deeper analysis with more functions embedded in Voyant.

In order to better answer the research question we proposed, apart from Voyant assist analysis, some textual materials are applied as well. Lin Daiyu is a significant character in THE DREAM OF THE RED CHAMBER. Every word and action of hers is carefully designed by the author. Her personality and destiny play a crucial role in the novel(Jiang,2020). Therefore, the analytical results of tools should remain consistent with the author's intention in creating this character. From the early stages of her life, she faced a fragmented and desolate family environment. Losing her mother in childhood and her father in adolescence, she became dependent on others.

This profoundly impacted her life, altering its trajectory and shaping her sensitive and melancholic personality. In the Jia Family, Lin Daiyu dislikes the conventional constraints of a household and admires Baoyu's nonchalant attitude towards life. Afterward, she fell in love with this man, holding the wish to have a beautiful and passionate love story, whereas engaged in ceaseless struggles with her destiny in this intense love affair. She was highly sensitive, and her love for the man was both humble and heartbreaking. Confronted with Baochai, who eventually married Baoyu, she concealed her inner humility, went against her original beliefs, and resisted, ultimately hurting herself(Ailipiya,2020). What encircled her is the cruelty of real life, the death of her mother, the independent life living in the Jia family, and the mental stimulation received from feudal society's normalcy(Jiang,2020). Under these intricate life experiences and living environments, she stands out by her robust and rebellious spirit. She fights against fate, conventional norms, and the bad habits of feudal society. That is why she frequently expresses her unwillingness and tries to be courageous enough to reject those she feels abhorrent, which can be reflected in our analysis. Nevertheless, the pains received from her childhood and suffering from the independent life in the Jia family, where the thoughts of class and feudalism are deep-rooted in every member's mind, result in her sensitive and negative emotions toward life.

Overall, the success of portraying Lin Daiyu, differing from other characters's language styles, is associated with the author's careful character design through word usage and emotional expression, effectively conveying her rebellious spirit, unique

qualities, and the complex emotions she experiences in the face of societal and familial challenges, as revealed through the textual analysis.

REFERENCES

- Ailipiya, A. (2020). Analysis on the Image of Lin Daiyu in A Dream of Red Chamber. *Conference on Basic Education Reform and Innovation in the Internet Environment in 2020*, 328–329.
- Benmin, Y. (2011). *The Analysis of the Yangs' Cultural Translation Strategies in A Dream of Red Mansions*. Master's thesis, Huazhong University of Science & Technology, China.
- Cao Xueqin, Gao e, Zhang Jun, Nie shiqiao, Zhou Jibin, & Gong shuduo. (2014). *The dream of the red chamber*. BEIJING BOOK CO. INC.
- Jiang, J. (2020). A Brief Discussion on the Characterization of Lin Daiyu in “Dream of the Red Chamber.” *Chinese Basic Education Knowledge Library*, 22, 16–17.
- Lyu Qixiang. (1987). The spirit of flower, the embodiment of poetry - the cultural connotation and modeling characteristics of Lin Tai-yü's image. *Journal of The dream of the red chamber*, (3), 48-67.
- Moretti, F. (2011). Network Theory, Plot Analysis. *New Left Review*, 68, 80-102.
- Poupard, D. (1984). Literature Criticism from 1400 to 1800. New Jersey: Gale Research Company.
- Zhu Weiming. (1994). Two ways of existence of life - the images of Lin Tai-yü and Xue Baochai and their cultural significance. *Journal of The dream of the red chamber*, (1), 155-166.