

Innovation in Applied Linguistics

Volume 1 Issue 1 2025
ISSN 3068-7845

MASON PUBLISH GROUP

Innovation in Applied Linguistics

PUBLISHER

Journal of Innovation in Applied Linguistics is published by Mason Publish Group, Seattle, USA; editorialoffice@masonpublish.org. For submission instructions, subscription and all other information visit: <https://masonpublish.org/>

COPYRIGHT AND LICENSING

Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a Creative Commons Attribution-ShareAlike 4.0 International License that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.

DISCLAIMER

The Publisher, Associations and Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher, Associations and Editors.

ONLINE OPEN

Innovation in Applied Linguistics accepts articles for Open Access publication. Please visit <https://www.caatj.com> for further information about Online Open.

© 2025 by MASON PUBLISH GROUP

Editorial Team

Editor-in-chief

Wenzhong Zhu, Guangdong University of Foreign Studies, PR.China

Associate editors-in-chief

Chunlin Yao, Tianjin Chengjian University, PR.China

Yunxia Zhu , University of Queensland, Australia

Jakub M. Pawlik, Wroclaw University of Economics and Business, Poland

Xiaoqin Liu, Guangdong University of Finance, PR.China

Editorial board

Zhibin Lin, University of Durham, UK

Xin Wen, Sehan University, Republic of Korea

Muchun Wan, Beijing International Studies University, PR.China

Siying Cai, Southern Medical University, PR.China

Jieqing Pan, Guangdong University of Chinese Medicine, PR.China

Baoxia Zuo, Tianjin University of Finance and Economics Pearl River College, PR.China;

PhD of Education, Sehan University, College of Education, Republic of Korea

Xin Wen, Sehan University, College of Education, Republic of Korea; PhD of Korean Language Education, Seoul National University, College of Education, Republic of Korea

Yanhong Hu, Huazhong University of Technology and Science

Yongliang Wang, North China University of Water Resources and Electric Power

Zhien Guo, Beijing Language and Culture University, PR.China

Qinglong Du, University of Jinan, PR.China

Fei Li, China University of Geosciences, PR.China

Congcong Wang, Huanghuai University, PR.China

Zhisong Li, Guangdong Songshan Politechnic, PR.China

Yan Zhang, Tangshan University, PR.China

Innovation in Applied Linguistics

Volume1 Issue 1 2025

The linguistic positivity bias in English literature of different genres: A diachronic study across 700 years

Yaqian Shi 1

Study on Chinese Commercial Bank's CSR Efficiency and Sustainable Development through Corpus Analysis and DEA Model

Muchun Wan,Ruizhao Ma 14

Exploring Academic Discursive Competence in Chinese EFL Undergraduates Abstract Writing of Research Articles

Fengjie Zhou,Fe Li..... 32

Examination of the Prevalent Biases in Digital Composition of Chinese Characters

Shubin Li..... 50

A Corpus-Based Comparative Study on the Low-carbon Image Construction of Chinese and American Banks

Xiaoqin Liu,Yishan Zhang,Zequan Jiang 59

The linguistic positivity bias in English literature of different genres: A diachronic study across 700 years

Yaqian Shi*

¹School of Foreign Languages, Huazhong University of Science and Technology, Wuhan, China

*Corresponding author Email: shiyaqian0719@163.com

Abstract

Previous studies have found that human beings have a positive tendency, which is labelled as the linguistic positivity bias. Such a bias has been examined and confirmed in literary works such as children and youth literature. However, such literature works are written for children and adolescence and thus they are naturally more positive. It is necessary to examine the bias based on literature of other genres that are written for adults. In addition, it remains unexplored how the positivity bias has evolved over a long period of time and how the social and cultural factors interact with literature works to shape the bias. To address the aforementioned issues, the present study aims to examine the linguistic positivity bias in English literature across three genres (i.e., novels, plays, and poetries) and their temporal trend across 700 years. The results show that most of the literature works have a positive tendency and no significant difference is found among the three genres regarding the sentiment values, which further confirms the linguistic positivity bias in human languages. However, it is worth noting that the linguistic positivity bias displays a decreasing trend. Possible explanations are provided for such a trend from social and cultural perspectives.

Keywords: linguistic positivity bias; English literature; genres; sentiment analysis

1 Introduction

The linguistic positivity bias, also known as the Pollyanna hypothesis (Augustine et al., 2011) or the Pollyanna principle (Armstrong & Hogg, 2001), refers to a universal human tendency to use a wide range of positive words more frequently (Dodds et al., 2015). One possible reason for the positivity bias is that human beings are generally optimistic and prefer to look at the bright or positive side of life (Boucher & Osgood, 1969). Another possible reason is that positively biased utterances can facilitate social relationships and help individuals achieve their communicative purposes (Garcia et al., 2012). As a result, they may unconsciously use more positive words in communications to convey their positive feelings such as happiness (Kloumann et al., 2012).

The linguistic positivity bias in human languages have been examined in various contexts (Augustine et al., 2011; Cao et al., 2021; Wen & Lei, 2021; Yuan et al., 2025). One area of research is to investigate the linguistic positivity bias across different languages (Augustine et al., 2011; Boucher & Osgood, 1969; Garcia et al., 2012). For example, Armstrong (2001) investigated the linguistic positivity bias in French. The result showed that young speakers of French tended to create more positive words while the use of new negative words remained stable, which validated the Pollyanna hypothesis. Similarly, the linguistic positivity bias has been examined and confirmed in other languages such as English, German, and Spanish (Augustine et al., 2011; Garcia et al., 2012), which further lend support to the universality of the linguistic positivity bias in human languages.

Another area of research is to investigate the linguistic positivity bias from the cognitive perspective (Jacobs, 2015; Lüdtke & Jacobs, 2015). For example, Lüdtke and Jacobs (2015) compared the processing speed of positive and negative words based on the word recognition task. Thirty-six native adult Germans were recruited to complete the task. They found that positive words have faster response times than neutral or negative ones, which indicated that positive words are cognitively easier to process. The same result is also obtained in 6–12-year-old children (Sylvester et al., 2016), which further confirmed the findings made in previous studies. The result can be explained by the higher subjective exposure frequency of positive words, which leads positive words to be processed much faster since they are more internally activated in memory (Jacobs et al., 2020).

Still another line of research focuses on the linguistic positivity bias across different genres such as academic writing and literature (Cao et al., 2021; Jacobs et al., 2020; Vinkers et al., 2015; Wen & Lei, 2021). For example, Vinkers et al. (2015) explored the use of positive and negative words in the abstracts of medical articles published between 1974 and 2014. Their analysis revealed an 880% rise in positive words and a 257% increase in negative words over the last 40 years. The result showed that researchers tended to use more positive words in their abstracts to promote their research, which demonstrated the linguistic positivity bias in academic writing. As a follow-up study, Wen and Lei (2021) examined the abstracts of 12 research disciplines in life sciences based on both lexicon analysis and sentiment analysis. The results showed a higher frequency of positive words in abstracts based on the lexicon analysis and a positive sentiment based on the sentiment analysis. In addition, the results also showed a significant upward trend of linguistic positivity bias over the past five decades, which further confirmed the linguistic positivity bias in academic writing.

More recently, researchers shifted their attention to linguistic positivity bias in literature (Dodds et al., 2015; Jacobs et al., 2020). For example, Dodds et al. (2015) examined the linguistic positivity bias in several languages such as English, French, Spanish, and Chinese based on 24 corpora, one of which is Google Books Project that includes many world-famous literature works. They investigated the use of positive and negative words in literature with a special word list. They found that positive words were significantly more frequent than negative ones in those literature under investigation, which demonstrated a positivity bias in literature. However, one possible issue of Dodds et al.'s (2015) study is its limited number of words in the word list, which includes only the most frequent 5–10,000 words. Therefore, their findings may not be con-



clusive. To address the issue, Jacobs et al. (2020) examined the linguistic positivity bias in literature based on the technique of sentiment analysis. To be more specific, they calculated the sentiment values of the English and German children and youth literature from the Gutenberg Literary English Corpus and the child-Lex corpus. Results of their study showed a positivity trend in those literature works, which added further evidence to the universality of the linguistic positivity bias.

Although the studies mentioned above have greatly enhanced our understanding of linguistic positivity bias in literature, their results remain incomplete and need further exploration. One issue is that most of the previous studies have not considered the factor of genres when examining the linguistic positivity bias in literature. Some studies such as Dodds et al. (2015) employed mixed genres of literature, which may confound the results. Other studies such as Jacobs et al. (2020) examined only one genre of literature, i.e., children and youth literature. Such literature works are written for children and adolescents who are in a critical period of value formation. It follows that they often convey positive information and conforms to the linguistic positivity bias. Therefore, the findings based on children and youth literature reveal only the features of children and youth literature and may not be generalized to literature of other genres. To paint a fuller picture of the linguistic positivity bias in literature, it is necessary to examine literature of other genres such as poetries and novels that are written for adults.

Another issue is that few of previous studies has examined the diachronic development of the linguistic positivity bias over a long period of time. Although previous studies have investigated the temporal trend of the positivity bias in academic writings, it remains underexplored how such a bias develops in other genres such as literature. Literature works are products of the age, which are significantly influenced by the social and cultural environments in which they were written. Therefore, it is interesting to diachronically investigate the trend of the linguistic positivity bias in literature based on a large dataset over a long period of time and explore how it interacts with the social and cultural factors (Iliev, et al., 2016; Iliev & Smirnova, 2016).

Still another issue is the method used to examine the linguistic positivity bias. Most of the previous studies examined the positivity bias based on a limited set of positive and negative words (except a few studies such as Jacobs et al., 2020 and Wen & Lei, 2021). That is, they examined the normalised frequency of positive and negative words used in texts based on a small predefined lexicon (Cao et al., 2021; Dodds et al., 2015). For example, Vinkers et al. (2015) investigated only 50 positive and negative words. The more frequent use of positive words indicates a positive sentiment and thus a linguistic positivity bias in the texts. However, this method is not a robust approach since it only includes a small number of words. In addition, they did not consider the intensity of positivity or negativity. For example, “bad” and “notorious” are negative words, but the latter word is obviously more negative than the former one. As a result, the findings may not be conclusive enough. Therefore, more sophisticated approaches are required to offer stronger and more reliable evidence for the linguistic positivity bias in human languages.

To address the aforementioned issues in previous studies, the present study aims to investigate the linguistic positivity bias in literature of different genres (i.e., novels, plays, and poetries) and their diachronic trend based on the new technique of sentiment analysis. More specifically, the purpose of the study is two-fold. First, it intends to calculate the sentiments of the literature of three genres, i.e., novels, plays, and poetries, and examine whether the linguistic positivity bias still holds true for them. Second, it aims to explore the diachronic trend of the linguistic positivity bias in literature over a period time of approximately 700 years (from 1321 to 20th century).

2 Sentiment Analysis

Recently, a new emerging method for the examination of linguistic positivity bias is sentiment analysis (Jacobs et al., 2020; Wen & Lei, 2021). Sentiment analysis is a technique employed to evaluate the sentiment conveyed by a text, such as whether it is positive or negative (Mäntylä et al., 2018; Taboada, 2016). There are two primary methods for conducting sentiment analysis: machine learning-based approaches and lexicon-based approaches (Liu & Lei, 2018). The machine learning-based method is a classification approach that relies on a classifier trained with human-labelled sentiment datasets to identify the sentiment of a given sentence or text (D’Andrea et al., 2015). This approach has a better accuracy in determining the sentiment of a sentence or a text. However, it has a limited applicability across different domains since sentiment-labelled data are not readily available. The lexicon-based approach calculates the sentiment values or scores of a text based on a large lexicon of sentiment words that have been manually annotated with sentiment scores (van Houtan et al., 2020). When the sentiment value or score of a text is larger than zero, its sentiment is positive. In contrast, when the sentiment value or score of a text is less than zero, its sentiment is negative. Compared to the machine learning-based approach, this approach has a wider application in texts of various genres and domains since the lexicon of sentiment words is field or domain neutral (Taboada, 2016).

Currently, some sentiment analysis tools have been developed for sentiment analysis, such as the packages of *syuzhet* (Jockers, 2017) and *sentimentr* (Rinker, 2021) in programming language R. They are developed based on many sentiment lexicons that consider the intensity of positivity and negativity. For example, the lexicon of *syuzhet* includes more than 10,000 commonly used words and their sentiment values vary from -1 to 1. A positive value means a positive emotion while a negative value indicates a negative emotion. The larger the absolute sentiment value of a word is, the more positive/negative the word is. Due to its large lexicon and wide applicability, *syuzhet* has been employed in various lines of research regarding sentiment analysis (Liu & Lei, 2018; Vergeer, 2020). However, one possible concern in *syuzhet* is that it does not consider valence shifters such as negators, which thus may lead to errors in the sentiment analysis of a sentence (Rinker, 2021). For example, the sentence of “He likes playing football” is positive and with a negator, the sentence “He doesn’t like playing football” becomes negative. Nevertheless, both sentences may be tagged as positive based on *syuzhet* since it fails to consider the negator, which makes the sentiment analysis less accurate. To address the concern, *sentimentr* that takes valence shifters into account has been developed as an alternative for sentiment analysis (Rinker, 2021), which can produce more accurate and valid results. Due to its satisfactory performance in sentiment analysis, *sentimentr* has been used in various domains (Idler et al., 2022; Ikoro et al., 2018; Roberts et al., 2022).

In this study, we will take the lexicon-based sentiment analysis for examining the linguistic positivity bias in literature due to its larger lexicon, domain neutral application, and more sophisticated algorithms. When conducting the sentiment analysis, we will adopt the package of *sentimentr* since it considers valence shifters and thus has a more satisfactory performance.

3 Methodology

In this section, we will first introduce the research questions that this study attempts to answer. Then, we will describe the data used in this study and the procedures of data processing and analyses.

3.1 Research questions

- (1) What is the sentiment in English literature of the three genres, i.e., novels, plays, and poetries? Do they conform to the linguistic positivity bias?
- (2) How has the linguistic positivity bias in English literature evolved in the past 700 years?

3.2 Corpus data

The data used in this study are English literature from The Corpus of the Canon of Western Literature (Green, 2017)¹. This corpus consists of many canonical literary works produced by authors from the western countries such as Italy, France, Great Britain, Germany, and United States. These canonical literature works have great aesthetic beauty and play an important role in shaping other literature, as well as western thoughts and cultures (Adler & Weismann, 2000; Leavis, 1948). Based on the cultural and stylistic features in each era, the canonical literature works are categorized into four chronological ages: (1) The theocratic age (2000 BCE to 1321 CE); (2) The aristocratic age (1321 CE to 1832 CE); (3) The democratic age (1832 CE to 1900 CE); (4) The chaotic age (20th Century). In addition, the literature works are tagged with genres such as fictions, plays, poetries, and mixed genres. The aforementioned features of The Corpus of the Canon of Western Literature make them particularly appropriate for the present study since 1) it contains different genres of literature, and 2) more importantly, it covers a long span of time, which allows for a diachronic study of the linguistic positivity bias.

Table 1 The statistics of English literature used in this study

Ages	Time span	Country/region	No. of texts	No. of words
The aristocratic age	1321 to 1832	England and Scotland	143	14,416,044
The democratic age	1832 to 1900	Great Britain	191	19,287,528
		United States	80	7,734,357
The chaotic age	20th century	Great Britain and Ireland	84	5,937,856
		Australia and New Zealand	3	212,723
		United States	41	1,889,639
Total	/	/	542	49,478,147

In this study, we only used the literature written in English (see Table 1 for the statistics). It should be noted that this corpus has no English literature at the theocratic age and thus the evolution of the linguistic positivity bias in this period is not investigated. That is, we explore the evolution of the linguistic positivity bias in literature across three ages (approximately 700 years), i.e., the aristocratic age, the democratic age, and the chaotic age.

The factor of genre may also confound the results since literature of different genres convey different emotions (Kim & Klinger, 2018). Therefore, we investigate the evolution of the linguistic positivity bias in English literature of three different genres, i.e., fictions, plays, and poetries (see Table 2 for the statistics). We choose the three genres for two reasons. First, they are main and common genres in literature that have a wide range of readers. Second, these three genres occupy the largest proportion in the corpus, which can secure a wide number of texts for the follow-up study. In addition, the consideration of genres can also ensure the validity and reliability of our results.

¹ <https://www.dropbox.com/s/xtv2r37ytf9pp7/Corpus%20of%20the%20Canon%20of%20Western%20Literature%20%281.0%29.rar?dl=0>

Table 2 The statistics of English literature of four genres, i.e., fictions, plays, and poetries

Ages	Fictions		Plays		Poetries	
	No. of texts	No. of words	No. of texts	No. of words	No. of texts	No. of words
The aristocratic age (1321 to 1832)	27	3452368	31	1354638	24	1838910
The democratic age (1832 to 1900)	112	15387092	8	293861	71	4459031
The chaotic age (20th century)	64	6636013	11	286525	34	389027
Total	203	25880849	50	1935024	129	6686968

3.3 Data processing and analyses

The English literature in The Corpus of the Canon of Western Literature were processed and analysed as follows.

First, the texts were cleaned with information such as Gutenberg terms of use, license, and notes deleted.

Second, the linguistic positivity bias in English literature was examined across the three genres (i.e., fictions, plays, and poetries). That is, we calculated the sentiment values of texts of English literature across the three genres based on the sentiment analysis. Then, outliers were detected. A text was considered an outlier and excluded from further analysis if its sentiment value was deviated by three standard deviations from the mean (Shiffler, 1988). Two outliers within the poetry genre were identified and removed prior to further analysis. Subsequently, a one-way ANOVA was performed to determine whether there were significant differences among the three genres. In cases where significant results were found, Games–Howell post hoc tests were applied to assess differences between paired genres.

Finally, the linguistic positivity bias in English literature was analysed across three historical periods. Sentiment values for the literature from these periods were calculated with sentiment analysis techniques. After identifying and removing three outliers, the remaining texts were used for further analysis. A one-way ANOVA was then conducted to assess whether significant differences existed among the three periods. If significant results were found, Games–Howell post hoc tests were applied to explore differences between paired periods.

The work mentioned above was completed with homemade scripts in the programming language R. The sentiment analysis was conducted based on the R package of sentimentr (Rinker, 2021).

4 Results

In this section, we will first report the sentiment results of English literature across the three genres and then present the results across the three ages.

4.1 Sentiment analysis results of English literature across the three genres

The descriptive statistics of the sentiment values of English literature in the three genres are shown in Table 3 and Figure 1. The results show that most of the literature in the three genres have a positive sentiment since most of their sentiment values are larger than zero. In addition, the result of one-way ANOVA analysis

shows no significant difference among the three genres ($df = 2$, $F = 0.907$, $p = 0.405$). That is, the sentiment values of the three genres have no significant difference.

Table 3 The descriptive statistics of the sentiment values of English literature across the three genres

Genres	Max	Min	Mean	S.D.
Fiction	0.160	-0.097	0.038	0.043
Poetry	0.130	-0.050	0.047	0.047
Play	0.234	-0.188	0.037	0.058

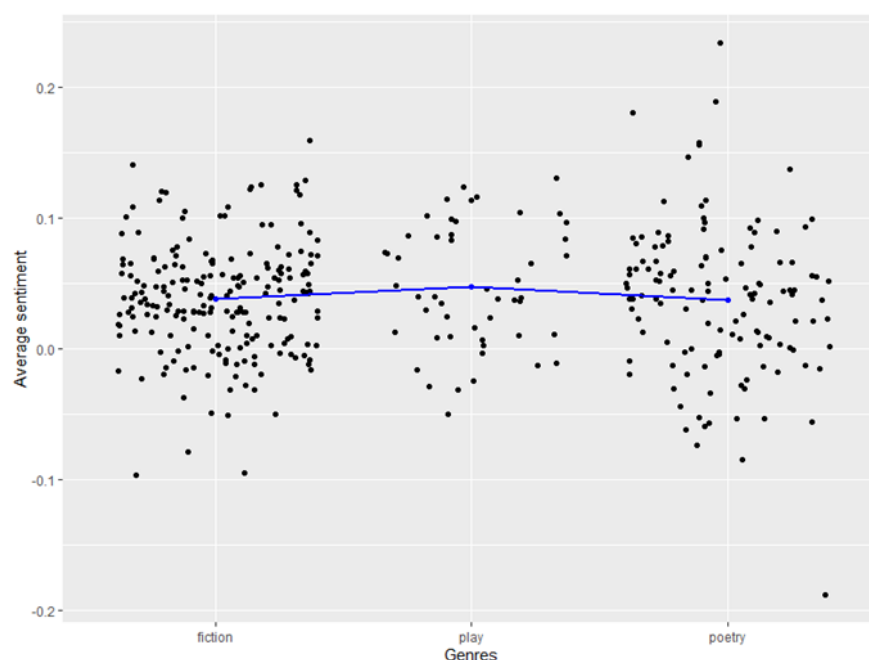


Figure 1 Sentiment results across the three genres (i.e., fictions, plays, and poetries)

4.2 Sentiment analysis results of English literature across 700 years

The descriptive statistics of the sentiment values of English literature across the three ages are shown in Table 4 and Figure 2. Two points are worth noting regarding the results. First, most of the literature works have a positive sentiment in the three ages since most of their sentiment values are larger than zero. For example, there are 143 literature works at the aristocratic age, of which 134 have a positive sentiment (i.e., their sentiment values are larger than zero) and nine have a negative sentiment (i.e., their sentiment values are less than zero). In addition, as shown in Table 4, the mean sentiment values of English literature are all larger than zero in the three ages. Second, the sentiment values of the literature display a decreasing trend over the three ages. For example, the mean sentiment value of the literature at the aristocratic age is 0.061 while those at the democratic and chaotic ages are 0.052 and 0.015, respectively.

Table 4 The descriptive statistics of the sentiment values of English literature across the three periods

Periods	Max	Min	Mean	S.D.
The aristocratic age	0.187	-0.097	0.061	0.045
The democratic age	0.189	-0.057	0.052	0.049
The chaotic age	0.158	-0.095	0.015	0.044

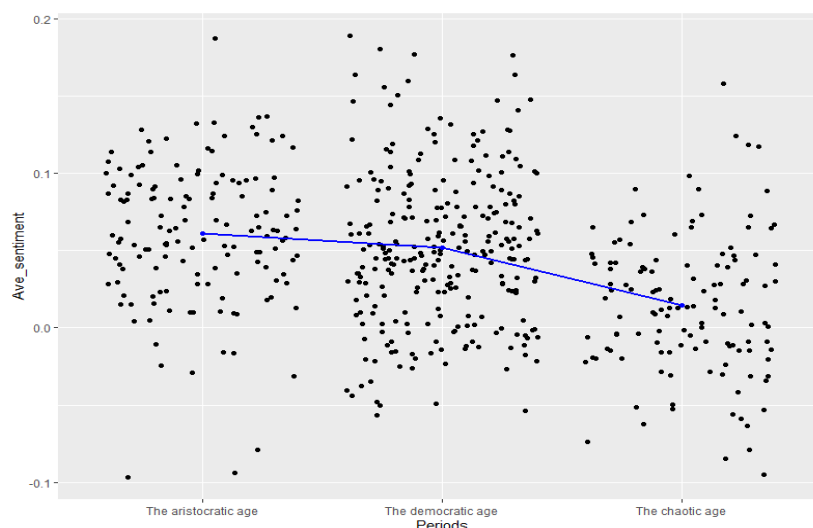


Figure 2 Sentiment values of the English literature across the three periods

The results of one-way ANOVA analysis show significant differences among the three ages ($df=2$, $F=37.37$, $p < 0.001$). Results of the post hoc test are shown in Table 5. The results indicate no significant difference between the democratic age and the aristocratic age. However, significant differences are found between the chaotic age and the aristocratic age/the democratic age. That is, the sentiment values are significantly decreasing from the democratic age to the chaotic age.

Table 5 The results of the post hoc analysis of the English literature across the three ages

Paired-age comparisons	p value
The democratic age vs. The aristocratic age	0.140
The chaotic age-The aristocratic age	<0.001
The chaotic age-The democratic age	<0.001

The descriptive statistics of the sentiment values of English literature across the three genres in the three ages are shown in Table 6 and Figure 3. As shown in Table 6, the mean sentiment values of two literature genres, i.e., fictions and plays, are also larger than zero. Interestingly, the mean sentiment value of poetries is approximately zero at the chaotic age. The mean sentiment values of all the three genres also display a decreasing trend over the three ages.

The results of one-way ANOVA analysis show that significant differences exist among the three periods in the genres of fictions ($df = 2$, $F = 12.76$, $p < 0.001$) and poetries ($df = 2$, $F = 18.33$, $p < 0.001$) while no significant difference is found in the genre of plays ($df = 2$, $F = 2.227$, $p = 0.119$). The results of the post hoc test on the fictions and poetries between paired ages are shown in Table 7. The results indicate that the sentiment values of fictions and poetries are significantly decreasing from the democratic age to the chaotic age.

Table 6 The descriptive statistics of the English literature of the three genres across the three periods

	fictions				plays				poetries			
	Max	Min	Mean	S.D.	Max	Min	Mean	S.D.	Max	Min	Mean	S.D.
The aristocratic age	0.124	-0.097	0.053	0.044	0.130	-0.031	0.057	0.050	0.137	0.011	0.071	0.030
The democratic age	0.160	-0.050	0.046	0.042	0.085	-0.011	0.041	0.035	0.189	-0.057	0.044	0.048
The chaotic age	0.088	-0.095	0.017	0.035	0.073	-0.050	0.024	0.039	0.158	-0.085	0.000	0.051
Total	0.160	-0.097	0.080	0.043	0.130	-0.050	0.047	0.047	0.234	-0.188	0.037	0.058



Figure 3 The diachronic trend of the sentiment values of the literature across the three genres in the three periods

Table 7 Results of the post hoc test on the sentiment values of the fictions and poetries between paired periods

Periods for comparison	p-value	
	fictions	poetries
The democratic age vs. The aristocratic age	0.688	0.05
The chaotic age vs. The aristocratic age	<0.001	<0.001
The chaotic age vs. The democratic age	<0.001	<0.001

5 Discussion

This study has examined the sentiments of English literature of three genres (i.e., fictions, plays, and poetries) and investigated their diachronic trend over a period of more than 700 years. The results reveal some points of interest.

First, the results reveal that the three genres show no significant difference in their sentiment values and most of them have a positive sentiment. The results confirm the linguistic positivity bias in these three genres, i.e., fictions, plays, and poetries, which are complementary to findings made in previous studies such as Jacobs et al. (2020). Jacobs et al. (2020) investigated the linguistic positivity bias in English and German children and youth literature works. Our study also found such a bias in other genres such as fictions, plays, and poetries. It indicates that the authors of different genres all tend to convey a positive feeling to readers. The result provides more evidence to the universality of the linguistic positivity bias in human languages.

Second, the results show that most of literature works have a positive sentiment in the three ages. The results also indicate a linguistic positivity bias in English literature, which is in line with the findings made in previous studies such as Jacobs et al. (2020) and Green (2017). For example, Green (2017, p. 289) analysed the most frequent lemmas used in the Corpus of the Canon of Western Literature and found that many of them were positive words such as good, great, and love. Based on the results, he concluded that the canon-

ical literature also shows a linguistic positivity bias despite authors from Homer to Hemmingway mainly address death, war, heartache, and tragedy. Our results further confirm his findings based on the new technique of sentiment analysis.

Third, the results reveal a downward trend of linguistic positivity in English literature over the three ages. In particular, the sentiment values of English literature at the chaotic age are significantly lower than those at the aristocratic and democratic ages. The results do not conform to those found in other genres such as academic writings. For example, Wen and Lei (2021) examined the diachronic trend of the linguistic positivity bias in academic writings across 50 years and found a significant upward trend. The main reason for such an inconsistency is attributed to the data type used in these two studies. The data used in their study are academic writings. One of the purposes of academic writing is to “sell” the research to the journals. Therefore, writers need to use some linguistic devices such as the use of positive words to promote their research.

In contrast, the data used in our study are literature works. One of the purposes of literature works is to express the writers’ thoughts, convey their emotions, and more importantly show the real world at their time (Fialho, 2019; Milner, 2005; Nurhamidah, 2019). That is, literary works are closely related with the social conditions in which those literary works are created. The aristocratic age (from 1321 to 1832) involves the Renaissance and Reformation period, the Enlightenment period, and the Romantic period, during which writers aimed to recover the philosophy in ancient Greece and Rome (Payne, 2014), emphasize reason and science (Hankins, 1985; Pinker, 2018), and focus on nature, imagination, and individuality (Izenberg, 1992). Therefore, during these periods, humanism, nature, and stories from classical mythology became the topics of literary creation. The democratic age (from 1832 to 1900) involves the Victorian period, during which writers longed for the morality of the medieval world (Timko, 1975). On the contrary, the chaotic age (the 20th century) involves the Modern Period in which the World Wars broke out. The society experienced enormous changes, such as the collapse of social order, the loss of faith in rational theory, and lack of authority and religion (Altaf, 2022). In this period, writers focused more on the inner feelings due to the disillusionment with the World Wars (Altaf, 2022; Yousef, 2017). As a result, the literary works in this period were more negative since individuals were more depressed and pessimistic due to the chaos and unrest in society (Van Hulle, 2018). For example, Eliot’s plays focused on moral and religious issues and entertained the audience with shrewd social satire (Lehman, 2009), which resulted in a negative sentiment in his works. It follows that the average sentiment values at the chaotic age are decreasing, which are significantly lower than those at the aristocratic and democratic ages.

This study has some theoretical implications for the hypothesis of the linguistic positivity bias. Our results indicate a more complex mechanism behind the linguistic positivity bias. It has been hypothesized that human beings tend to use more positive words since they often look at the bright side of the life (Augustine et al., 2011). That is, the linguistic positivity bias is driven by our cognition. Based on this hypothesis, the linguistic positivity bias should present an increasing trend or remain the similar level. However, our results show a decreasing trend of the bias over the past 700 years. This finding may reveal that other factors such as cultural and social ones also contribute to the linguistic positivity bias (Iliev, et al., 2016). That is, the linguistic positivity bias may be shaped by both the internal factor such as our cognition and the external factors such as social and cultural ones. Therefore, it is necessary to consider mixed factors when researchers investigate the linguistic positivity bias and explore their complex relationships behind such a bias.

6 Conclusion

This study examined the linguistic positivity bias in literature of three genres (i.e., novels, plays, and poetries) and explored their diachronic trend across 700 years. The results show that most of the literature works have a positive sentiment and no significant difference is found among the three genres, which further confirm the linguistic positivity bias in the field of literature. However, it is worth noting that the linguistic positivity bias displays a decreasing trend. This may be attributed to the social and cultural factors. The finding indicates a more complex mechanism of linguistic positivity bias. That is, the linguistic positive



bias may be not only driven by our cognition but also shaped by other external factors such as social and cultural ones.

This study may be limited in the following aspects. First, it examined only three genres. Future research may examine the linguistic positivity bias based on other genres of literature such as prose and drama since different genres may convey different emotions (Kim & Klinger, 2018). Second, this study investigated the temporal trend of the linguistic positivity bias based on literature. The findings may not be generalized to other genres or registers. Future studies may diachronically examine the bias based on other written corpora such as the Corpus of Historical American English since it includes other genres such as newspaper and magazines. In addition, researchers may also explore its temporal trend based on the spoken corpora since they may better reflect the feelings or emotions of the general public at that time. Finally, the literature used in this study only cover the first half of the 20th century. Future research may investigate the temporal trend of the linguistic positivity bias in recent decades since the society has changed significantly due to the rapid development of science and technology. It is of interest to investigate how the cognitive factor and social factors interact with each other to shape the trend of the linguistic positivity bias in recent decades.

Funding

Fundamental Research Funds for the Central Universities (Grant Number: 2025WKQN006).

Conflict of interest disclosure

The author(s) declared that there are no conflicts of interest with respect to the authorship or the publication of this article

References

- Adler, M. J., & Weismann, M. (2000). *How to think about the great ideas: From the great books of western civilization*. Chicago: Open Court.
- Altaf, H. (2022). Modernism and Disillusionment: A Study of the Prominent Works of the British Modern Period. *The Journal for English Language and Literary Studies*, 12(3), 10-10.
- Armstrong, N., & Hogg, C. (2001). The Pollyanna Principle in French: A study of variable lexis. *Journal of Pragmatics*, 33(11), 1757–1785. [https://doi.org/10.1016/S0378-2166\(00\)00081-3](https://doi.org/10.1016/S0378-2166(00)00081-3)
- Augustine, A. A., Mehl, M. R., & Larsen, R. J. (2011). A positivity bias in written and spoken English and its moderation by personality and gender. *Social Psychological and Personality Science*, 2(5), 508–515.
- Bloom, H. (1994). *The western canon: The books and school of the ages*. Boston: Houghton Mifflin Harcourt.
- Boucher, J., & Osgood, C. E. (1969). The pollyanna hypothesis. *Journal of Verbal Learning and Verb Behavior*, 8, 1–8.
- Cao, X., Lei, L., & Wen, J. (2021). Promoting science with linguistic devices: A large-scale study of positive and negative words in academic writing. *Learned Publishing*, 34(2), 82–88.
- D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 26–33.
- Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Mitchell, L., Harris, K. D., Kloumann, I. M., Bagrow, J. P., Megerdumian, K., McMahon, M. T., Tivnan, B. F., & Danforth, C. M. (2015). Human language reveals a universal positivity bias. In *Proceedings of the National Academy of Sciences of the United States of America*, 112(8), 2389–2394.
- Fialho, O. (2019). What is literature for? The role of transformative reading. *Cogent Arts & Humanities*, 6(1), 1692532.

- Garcia, D., Garas, A., & Schweitzer, F. (2012). Positive words carry less information than negative words. *EPJ Data Science*, 1(3), 1–12.
- Green, C. (2017). Introducing the Corpus of the Canon of Western Literature: A corpus for culturomics and stylistics. *Language and Literature*, 26(4), 282–299. <https://doi.org/10.1177/0963947017718996>
- Hankins, T. L. (1985). *Science and the Enlightenment*. Cambridge: Cambridge University Press.
- Idler, E., Bernau, J. A., & Zaras, D. (2022). Narratives and counter-narratives in religious responses to COVID-19: A computational text analysis. *PloS one*, 17(2), e0262905.
- Ikoro, V., Sharmina, M., Malik, K., & Batista-Navarro, R. (2018, October). Analyzing sentiments expressed on Twitter by UK energy company consumers. In *2018 Fifth international conference on social networks analysis, management and security (SNAMS)* (pp. 95-98). IEEE.
- Iliev, R., Hoover, J., Dehghani, M., & Axelrod, R. (2016). Linguistic positivity in historical texts reflects dynamic environmental and psychological factors. *Proceedings of the National Academy of Sciences*, 113(49), E7871-E7879.
- Iliev, R., & Smirnova, A. (2016). Revealing word order: Using serial position in binomials to predict properties of the speaker. *Journal of Psycholinguistic Research*, 45, 205-235.
- Izenberg, G. N. (1992). *Impossible Individuality: Romanticism, Revolution, and the Origins of Modern Selfhood, 1787-1802*. Princeton: Princeton University Press.
- Jacobs, A. M. (2015). Neurocognitive poetics: Methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, 9, 186.
- Jacobs, A. M., Herrmann, B., Lauer, G., Lüdtkke, J., & Schroeder, S. (2020). Sentiment Analysis of Children and Youth Literature: Is There a Pollyanna Effect? *Frontiers in Psychology*, 11, 574746.
- Jockers, M. L. M. (2017). Syuzhet: Extracts sentiment and sentiment-derived plot arcs from text. <https://CRAN.R-project.org/package=syuzhet>
- Kim, E., & Klinger, R. (2018). A survey on sentiment and emotion analysis for computational literary studies. *arXiv preprint arXiv:1808.03137*.
- Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., & Dodds, P. S. (2012). Positivity of the English language. *PloS One*, 7(1), e29484.
- Leavis, F. R. (1948). *The great tradition: George Eliot, Henry James, Joseph Conrad*. London: Chatto & Windus.
- Lehman, R. S. (2009). Eliot's Last Laugh: The Dissolution of Satire in "The Waste Land". *Journal of Modern Literature*, 65-79.
- Liu, D., & Lei, L. (2018). The appeal to political sentiment: An analysis of Donald Trump's and Hillary Clinton's speech themes and discourse strategies in the 2016 US presidential election. *Discourse, Context & Media*, 25, 143–152.
- Lüdtkke, J., & Jacobs, A. M. (2015). The emotion potential of simple sentences: Additive or interactive effects of nouns and adjectives? *Frontiers in Psychology*, 6, 1137.
- Mäntylä, M. V., Graziotin, D., & Kuuttila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16–32.
- Milner, A. (2005). *Literature, culture and society*. New York: Psychology Press.
- Nurhamidah, I., Purwanto, S., & Ekaningsih, N. (2019). A literary work as self-reflection of the author: Why and how it is manifested. *EduLite: Journal of English Education, Literature and Culture*, 4(2), 194-203.
- Payne, M. (2014). Renaissance or Reformation?. *A Companion to British Literature*, 70-81.
- Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress*. London: Penguin UK.
- Rinker, T. (2021). Calculate text polarity sentiment. <https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf>



- Roberts, A., Jellicoe, M., & Fox, K. (2022). How does a move towards a coaching approach impact the delivery of written feedback in undergraduate clinical education?. *Advances in Health Sciences Education*, 1-15.
- Shiffler, R. E. (1988). Maximum Z scores and outliers. *The American Statistician*, 42, 79–80.
- Sylvester, T., Braun, M., Schmidtke, D., & Jacobs, A. M. (2016). The Berlin Affective Word List for Children (kidBAWL): Exploring Processing of Affective Lexical Semantics in the Visual and Auditory Modalities. *Frontiers in Psychology*, 7, 969. <https://doi.org/10.3389/fpsyg.2016.00969>
- Taboada, M. (2016). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*, 2(1), 325–347.
- Timko, M. (1975). The Victorianism of Victorian Literature. *New Literary History*, 6(3), 607-627.
- van Houtan, K. S., Gagne, T., Jenkins, C. N., & Joppa, L. (2020). Sentiment analysis of conservation studies captures successes of species reintroductions. *Patterns*, 1(1), 100005. <https://doi.org/10.1016/j.patter.2020.100005>
- Van Hulle, D. (2018). Introduction Negative Modernism: Beckett's Poetics of Pejorism and Literary Enactment. *Beckett and Modernism*, 1-18.
- Vergeer, M. (2020). Artificial intelligence in the Dutch press: An analysis of topics and trends. *Communication Studies*, 71(3), 373–392. <https://doi.org/10.1080/10510974.2020.1733038>
- Vinkers, C. H., Tjldink, J. K., & Otte, W. M. (2015). Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis. *Clinical Research*, 351, h6467. <https://doi.org/10.1136/bmj.h6467>
- Wen, J., & Lei, L. (2021). Linguistic positivity bias in academic writing: A large-scale diachronic study in life sciences across 50 years. *Applied Linguistics*. Advance online publication. <https://doi.org/10.1093/applin/amab037>
- Yousef, T. (2017). Modernism, postmodernism, and metamodernism: A critique. *International Journal of Language and Literature*, 5(1), 33-43.
- Yuan, Z. M., Jiang, C., & Yao, M. (2025). The art of positive reporting in media discourse: Time magazine's linguistic positivity bias over the 20th century. *Digital Scholarship in the Humanities*, fqaf039.

Study on Chinese Commercial Bank's CSR Efficiency and Sustainable Development through Corpus Analysis and DEA Model

Muchun Wan*, Ruizhao Ma

School of English Studies, Beijing International Studies University, Beijing, 100024, China, wanmuchun@bisu.edu.cn

School of English Studies, Beijing International Studies University, Beijing, 100024, China, 2024110116@stu.bisu.edu.cn

*Corresponding author Email: 2024110116@stu.bisu.edu.cn

Abstract

By applying the efficiency DEA model and corpus analysis tool Antconc into the analysis of Corporate Social Responsibility report (CSR report) and taking commercial banks' performance of CSR as an input which may bring banks effective output, this research investigates CSR input-output efficiency and the achievement of sustainable development of China's major commercial banks by means of qualitative, quantitative and comparative analysis on the basis of the outcome of DEA empirical study. DEA model finds that the CSR input efficiency of 6 smaller commercial banks including China Citic Bank (CITIC Bank), Huaxia Bank (HB), China Guangfa Bank (CGB), China Merchants Bank (CMB), China Industrial Bank (CIB) and China Bohai Bank (CBB); the other 9 commercial banks including the 4 big banks in China are the inefficient banks. Then 30 keywords and 6 semantic domains of DEA efficient banks' CSR reports are summarized relative to DEA inefficient banks' through corpus tool's keyword analysis. Based on discursive context of CSR reports and characteristics of commercial banks sustainable development, this research finds that DEA efficient banks achieve a better sustainable development than DEA inefficient banks in corporate culture, stakeholder concern, marketization level, application of Fintech, green finance and business innovation, reference to exterior supervision mechanism. In the end, several advices are given for China commercial banks to improve their CSR input efficiency and achieve sustainable development. Different from the previous single linguistic perspective of CSR report, this study uses corpus tool for text analysis on the basis of empirical research on economic efficiency model. The combination of the two can avoid the subjectivity of traditional linguistic research, making the results of text analysis can more effectively serve the improvement of CSR input efficiency of commercial banks in order to achieve better sustainable development. At the same time, this research also enriches the study of business discourse analysis such as CSR reports.

Keywords: CSR input; efficiency; DEA; sustainable development; commercial banks



1 Introduction

As China has gradually entered the transition period, with the fast development of the economy, the outbreak of many social issues has also aroused strong public concern. Corporate social responsibility, CSR for short, has gradually become a topic of public concern and discussion. Michael Porter, the father of the competition strategy, once pointed out that corporate social responsibility is not just a cost but also a source of opportunity, innovation and competitiveness (Porter, 2006). The investment in CSR is not only a manifestation of corporate social responsibility, but also a strategic choice for enterprises to enhance their economic and social benefits, so it becomes an inevitable choice for corporate development. The improvement of CSR not only relies on the input of the company in social responsibility, but also depends on whether the company can reach its expected goal, that is, the efficiency of the CSR input.

Nowadays, the study of corporate social responsibility issues in the coal, food, and pharmaceutical industry has been relatively in-depth, but the study of corporate social responsibility in the commercial banks is still in its infancy. As the most important supplier of social financing under the background of supply-side structural reform, the commercial banks provide the core strength for China's social stability and sustainable development of China's economy. Compared with general enterprises, the fulfillment of the commercial banks' CSR has a radiation effect. The commercial banks can affect their customers' and related industries' CSR performance through green credit policy and investment concepts. In addition, while performing CSR, the commercial banks can realize the optimal allocation of social financial resources, the control of risks, the protection of environment and the reduction of energy consumption through the capital's leading function, thus achieving the sustainable development of economy and environment. Fundamentally, the commercial banks promote the harmonious development and mutual benefit among the banks themselves, the environment and the economy by fulfilling CSR, which directly or indirectly affects the economic stability and overall security of the entire country. For example, the commercial banks' credit supply for the real estate industry will affect the price of real estate assets, and the loosening level of the commercial banks' credit policy will also indirectly affect the consumer price index by affecting the amount of money supply. Therefore, the fulfillment of the commercial banks' CSR is not only related to their own sustainable development, but also conducive to the stable development of the entire economy and society. The study of the efficiency of the commercial banks' CSR input will enable commercial banks to change the stereotype that CSR input is just a no-output cost, increase their CSR input efficiency and operating efficiency, thus increasing the efficiency of their CSR input and actively performing their social responsibilities, such as lowering the financing cost in the green industry, the new-economy industry, the small and micro enterprises, and the inclusive finance area by proactively conducting business innovation. In general, the commercial banks can increase their soft power and gain public recognition by fulfilling their CSR while creating profits. Thus, commercial banks can realize the comprehensive sustainable development between themselves, the economy and the society in order to serve China's supply-demand reform better.

The Corporate Social Responsibility Report (CSR report) is a public report reflecting the self-discipline of the enterprise. It is mainly used to review the modern enterprise's service and contribution in the environment, consumers, employees, communities, shareholders and other public areas. (Bhatia, 2012; Zhong et al., 2015) As an important tool for communicating with stakeholders, CSR reports have become increasingly important. Textual analysis techniques that investigate the relationship between CSR performance and CSR disclosure have been used in several studies, with the specific techniques used ranging from counting the number of words in disclosures to creating indices to measuring the quality of entire CSR reports. For example, Wiseman's (1982) index, developed to analyze annual reports for determining the extent of CSR disclosure, is one of the first and most widely used indices in earlier research that evaluated environmental performance (e.g. Neu et al., 1998). Some studies find a positive relationship between CSR performance and CSR disclosure (e.g. Al-Tuwaijri et al., 2004; Clarkson et al., 2008; Luo and Tang, 2014). In view of the fact that CSR reports of China commercial banks are long and complex, content analysis concerning CSR reports have a certain degree of subjectivity (Su, 2010). A corpus-based approach to CSR report from a

linguistic point of view where information technology is used to allow the identification of routine patterns of lexical and co-textual distribution over large volumes of data (Mautner, 2005) pays more attention to the cumulative effect of language and reduces the subjective bias of researchers. The function of corpus tool's keyword list, concordance, collocation can provide objective and effective quantitative analysis tools for the comparative analysis of the text features of the CSR reports of the CSR efficient and inefficient commercial banks.

Thus by applying the efficiency DEA model and corpus analysis tool Antconc into the analysis of Corporate Social Responsibility report (CSR report), this research takes China major commercial banks' performance of CSR as an input which may bring banks effective output and objectively evaluate the CSR efficiency of each commercial banks through the DEA model; then on the basis of outcome of DEA model, by adopting the quantitative, qualitative, and comparative analysis, this research examines keyword features of DEA efficient commercial banks' CSR report relative to DEA inefficient banks' through Corpus retrieval tool Antconc. In addition, according to the assessment criteria of commercial banks' sustainable development in previous study and specific semantic context in CSR report, causes of the DEA efficient and inefficient banks are inferred and suggestions for China major commercial banks' improvement of CSR input-out efficiency and achievement of sustainable development are given in the end.

2 Literature Review

2.1 Research on Corporate social responsibility and sustainable development

Corporate social responsibility (CSR) and sustainable development (SD) are seen as the two of the main drivers for businesses including the banking industries over the past decade or so. CSR is becoming an ever more important field for business. Today's corporates ought to invest in CSR as part of their business strategy to become more competitive. CSR could be defined as the voluntary commitment of companies to contribute to sustainable economic development by integrating social and environmental concerns into their operations and interactions (Banerjee, 2008). Corporate success depends on the local environment including an appropriate infrastructure, the right types and quality of education to future employees, co-operation with local suppliers, quality of institutions and local legislation. In this corporate competitive context, the entrepreneurs' social initiatives can have great impact not only for themselves but also for the local society in terms of sustainable development (Porter, 2003).

Development can be defined as eliminating poverty, unemployment and inequality. Sustainability is associated with satisfying a national economy's needs for natural resources without compromising generations (Luke, 2013). Sustainable development is defined as "the development that meets the needs of present without comprising the ability of future generations to meet their own needs" (UN Brundtland Report, 1987). According to Bersler's model, there are three dimensions of sustainability: societal sustainability, economic sustainability, environmental sustainability (Besler, 2009). A business model, which has the value creation at the core, will allow concepts of CSR and sustainability because value creation is the primary motivator (Wheeler, Colbert & Freeman, 2003). It is impossible for a company to achieve its sustainable survival without creating value to its employees, shareholders, customers, society and all of its stakeholders because of the new dimensions of competition such as environmental politics (Iyiun, 2015). A company could not be successful by focusing on only its short-term based economic performance, neglecting environmental and societal performance (Besler, 2009). The overwhelming complexity of sustainable development has increased the goals for companies and entrepreneurs to take on responsibility for social and environmental challenges in a more proactive way and beyond mere legal compliance (Seelos & Mair, 2004). Besides, research has been conducted to examine the status of the research concerning CSR and sustainability in the management literature and finds that recently open calls for papers suggests that scholars will continue their efforts to shed more light on the two topics (Kudlak & Kathleen, 2015) Special issues on CSR / corporate sustainability also put special emphasis on how businesses organized and maintained their sustainability



orientation (see e.g., Journal of Business Ethics, 2009, 2012, 2013). A large number of literature shows that corporate's performance of CSR will lead to their sustainable development. The connotation of corporate's CSR and the principle of sustainable development are tightly connected and they extend the traditional fields of business ethnic jointly and enrich the theme of corporate governance.

As for the commercial bank---the foundation of the financial system, their social responsibilities aren't limited to the environmental profits making, resources reserving and philanthropy. Instead, the essential connotation of banks' CSR is to develop a series of proper mode of business operation and focus on stakeholder's concern since they are the allocator of financial resources and risk manager of the whole society. Commercial banks should combine CSR with bank' business, such as green finance and SMEs loan and find the commercial opportunity in the performance of social responsibility in order to achieve a sustainable win-win situation with society. Chen (2002) defined the sustainable development of commercial banks as commercial banks should guarantee the effective operation and steady development through refining internal business operation, taking the advantage of resource allocation and strengthening the capacity of development potentials which requires both the quantitative and qualitative financial development of commercial banks in order to achieve the overall financial efficiency of the society. Thus the research on commercial bank's improvement performance of CSR and achievement of sustainability which contains dual function will contribute not only banking industry but also other corporates' even the whole society's CSR performance and sustainable development.

2.2 Research on CSR and DEA model

Since Oliver Sheldon proposed the concept of CSR in the 1820s, there has been debate in the academic community about what CSR includes and whether companies should fulfill their social responsibilities. In recent years, the correlation between the social performance resulting from the fulfillment of corporate social responsibility and the operational performance has become a research hotspot. But so far, due to the inconsistency of indicators selection, sample selection methods and other control variables selection in the regression model that validates the correlation between the corporate social performance and the corporate operational performance, the correlation study between the two has not yet reached a unified conclusion. (Ding & Fu 2003) Porter & Kramer (2003), Schniez (2005), Yao Haixin et al (2005), Wang et al (2013) believe that there is a positive correlation between them, that is, fulfilling CSR can help improve the company's financial performance; Freeman (1999), Amir Barnea (2010) believe that the company's investment in fulfilling CSR is bound to waste capital and other resources, which indirectly affects the company's profits, that is, the two are negatively correlated. Margolis & Walsh analyze and summarize 127 documents and still cannot reach a unified conclusion on the relationship between the corporate social performance and the corporate operational performance. (Margolis & Walsh, 2007). This article believes, with the development of stakeholder theories and the rise of corporate citizenship, it has become the consensus of the academic community and the practice community that companies should fulfill their social responsibilities. Therefore, how to effectively fulfill social responsibilities has become today's new focus, and the research on the effectiveness of corporate social responsibility investment can provide theoretical guidance for enterprises to better conduct social responsibility investment and achieve sustainable development.

Charnes, Cooper, and Rhodes proposes a method for evaluating the efficiency of data envelopment analysis (DEA), and successfully uses the DEA method to evaluate the relative effectiveness of departments (Charnes et al., 1978). Su use the DEA analysis method to evaluate the listed company's corporate efficiency in the study of the relevance of CSR and corporate efficiency (Su et al., 2010). Ding and others also use the DEA method to evaluate corporate efficiency in the study of the impact of the social responsibility of large enterprises in China and the United States on business efficiency. (Ding et al., 2015) Although the above research on CSR has adopted the DEA method, it only considers the DEA value as an indicator to evaluate the operating efficiency of the enterprise rather than regard the implementation of CSR by the enterprise as an input indicator that can bring output.

Lin applies the DEA method to analyze the effectiveness of corporate CSR from the perspectives of input and output, and used compliance expenditure, employee expenditure, employee training expenditure, product assurance expenditure, environmental protection expenditure and charitable donation expenditure as the input indicators, and the main business income and main business profit as output indicators to conduct empirical research on the social corporate investment efficiency of listed companies in different industries, scales, and years, and proposed suggestions to help listed companies improve the efficiency of CSR investment (Lin, 2007). Song uses the DEA model, the book income tax, wage and welfare rate, earnings per share, interest cover ratio, main business cost rate, and accounts payable turnover rate as input indicators, and the total asset return rate and the Tobin Q as output indicators to analyze the CSR efficiency of the coal industry (Song, 2013). Bian et al. adopt the DEA-Malmquist model to analyze the investment efficiency of social responsibility of listed companies in the food manufacturing industry from both static and dynamic aspects, and propose relevant recommendations (Bian, 2015). From the existing literature, the academic community has attempted to use the DEA method to analyze the CSR efficiency from the input-output perspective, but there is still a certain degree of subjectivity in the process of quantifying input indicators. And from an overall perspective, the existing research is mainly focused on industries with high corporate social responsibility issues, such as excess capacity companies and food and pharmaceutical companies. As an important part of the current supply of funds in the process of economic transition, the commercial banks will be able to achieve harmonious development mutual benefit and long-term coexistence of their own, environment and economy. It will directly or indirectly affect the economic stability and the security situation of the entire country. The current research on corporate social responsibility of commercial banks is mainly focused on the research on the corporate social responsibility of commercial banks from the perspective of green finance (Sun, 2009), the research on corporate social responsibility evaluation standards and mechanisms of commercial banks (Zhu, 2008), and the relationship between corporate social responsibility and sustainable development of the commercial banks (Guo, 2009). However, it has not been fully valued by the academic and practical communities to focus on the efficiency of the corporate social responsibility performance of commercial banks and their promotion strategies from the perspective of efficiency. However, the study of efficiency of commercial banks' CSR performance of and their promotion strategies from the efficiency point of view have not been fully valued by the academic and practical communities.

2.3 Research on CSR, SD and corpus analysis

The Corporate Social Responsibility Report (also referred to as the Sustainability Report) refers to the disclosure of the social responsibility that the company undertakes, which mainly includes environmental, social, economic, stakeholder, and volunteer activities. (Dahlsrud, 2008)

The studies of CSR reports are mainly based on the perspective of economics and business management which adopt content analysis method to select indicators from CSR reports to study the relevance of CSR implementation and the company's operating efficiency as well as the concept of sustainable development connoted in company's CSR report. Some study analyses the relationship between the complexity of corporate social responsibility (CSR) disclosure and actual CSR performance by using several readability and disclosure size measures from computational linguistics, and finds that increased CSR disclosure and more readable CSR reports are associated with better CSR performance. (Nazari & Hrazdil, 2017) Li (2010) maintains that the textual analysis techniques developed in computational linguistics (e.g. Jurafsky et al., 2000) are useful tools for evaluating managerial intent. Cho et al. (2010) analyses the content of a small sample of U.S. companies to investigate if linguistic constructs suggestive of 'optimism' and 'certainty' used in CSR disclosure can provide a signal for actual CSR performance. Fuoli (2016) adopts corpus-based analysis to investigate how companies use stance expressions to construct and promote a positive corporate identity in their annual and CSR reports to gain the trust of the stakeholders. Carroll starts from the perspective of cultural differences and conclude that people in different countries have different perceptions of CSR which will cause different content preference when compiling CSR reports. (Carroll, 2004) From the per-



spective of linguistic genre, Bhatia proposed that in addition to the basic information disclosure function, CSR report also carries the function which help company to advocate its economic effectiveness and self correct negative news, so as to provide readers with a reference through reading the CSR report to perceive the corporate image(Bhatia, 2012). Fuoil introduced appraisal theory to compare CSR reports of BP and IKEA. Through evaluating the use of different vocabularies in the CSR report, he analyzed the information the two companies wanted to pass to the reader and found out the difference between the two in shaping their own corporate image(Fuoil, 2013). As for the research on corporate's concept of sustainability, some scholars have used close reading and other forms of qualitative analysis to analyze how "sustainable development" was constructed in a given text (Porter, 2005; Sandstrom, 2005). Huang and Zhu (2017) makes a study on Chinese corporations' conception of sustainable development based on an innovative view of corpus analysis, and finds out that Chinese companies should enhance their concept of sustainable development in business transactions.

All in all, the academic community has studied the CSR report from the perspective of linguistics and corpus, but there are still deficiencies in the number of studies and the depth of research. The existing research is mainly from the perspective of linguistics, using corpus tools to analyze the language features of CSR reports, summarizing the concept of sustainable development of the company, and providing suggestions for the writing of CSR reports and reading of stakeholders' CSR reports. It does not provide suggestions on how to improve company's CSR input efficiency and achieve sustainable development from the perspective of the combination of qualitative and textual analysis. Thus taking China major commercial bank as the research target, this paper firstly evaluates the CSR input efficiency of commercial banks and classify them into the CSR efficient bank and inefficient bank. Then the comparative textual analysis of CSR efficient banks' CSR report and CSR inefficient banks' through corpus tool are conducted and keyword features of DEA efficient commercial banks' CSR report relative to DEA inefficient banks' are examined through Corpus retrieval tool Antconc. In addition, according to the assessment criteria of commercial banks' sustainable development in previous study and specific semantic context in CSR report, causes of the DEA efficient and inefficient banks are inferred and suggestions for China major commercial banks' improvement of CSR input-out efficiency and achievement of sustainable development are given in the end. Different from the previous single linguistic perspective of CSR report, this study uses corpus tool for text analysis on the basis of empirical research on economic efficiency model. The combination of the two can avoid the subjectivity of traditional linguistic research, making the results of text analysis can more effectively serve the improvement of CSR investment efficiency of commercial banks in order to to achieve better sustainable development. At the same time, this research also enriches the study of business discourse analysis such as CSR reports.

3 Research Design

3.1 Research Questions

According to the review and analysis of previous literature, this study mainly examines 4 issues:

What are the DEA efficient and non-efficient banks respectively?

What are the salient features of the keyword list of DEA efficient banks' CSR reports reference to non-efficient banks' CSR reports?

Why do keyword list of DEA efficient banks' CSR reports have such feature?

How does China major commercial bank to improve their CSR input efficiency in order to achieve sustainable development better?

3.2 Research Tools

3.2.1 DEA Model

Performance measurement refers to the relationship between inputs and output. Evaluating the efficiency of organizational units is usually a difficult problem, especially when the multiplicity of inputs and outputs is required to be considered (Cook and Seiford, 2009; Wu, 2006). Data Envelopment Analysis (DEA) model is used in this paper to analyze commercial bank's multi-dimensional CSR performance measurement system in order to calculate the rank of CSR efficiency of each bank. The data envelopment analysis (DEA) is a nonparametric method to measure relative efficiencies of multiple decision-making units (DMUs), which has been recognized as a valuable analytical instrument and a practical decision support tool for dealing with the task of business performance assessment. The basic principle of DEA is to evaluate the relative effectiveness of multiple input, output -multiple decision-making units (DMUs)-commercial bank in this paper through the use of linear programming. Especially, DEA has no need to explicitly specify a mathematical form for the production function, and it can analyze and quantify the inefficiency of every DMU. DEA has been applied extensively to evaluate the performance for a variety of sectors including banks.

CCR model are used in this paper to examine the technical efficiency (TE) which is the comprehensive CSR input-output efficiency of commercial banks. The DEA CCR model (Charnes et al., 1978) assumed that there are n DMUs, with m inputs and s outputs, while the efficiency evaluation model of DMU can be defined as following: x_{ij} false represents the i th input value for j th DMU and the y_{rj} false represents the r th output value for j th DMU. Thus, the vector of the total input factor and the output element of the j th decision unit DMU_j false can be recorded as:

$$X_j = [x_{1j} \ x_{2j} \ \dots \ x_{mj}]^T, j = 1, 2, \dots, n \text{ false and } Y_j = [y_{1j} \ y_{2j} \ \dots \ y_{sj}]^T, j = 1, 2, \dots, n \text{ false}$$

The whole CCR model can be defined as following:

$$\begin{aligned} & \min \theta & (1) \\ & \sum_{j=1}^n X_j \lambda_j + s^- = \theta X_0 \\ s.t. \quad & \sum_{j=1}^n Y_j \lambda_j - s^+ = Y_0 & f \\ & \lambda_j \geq 0 \\ & s^+ \geq 0, s^- \geq 0 \\ & j = 1, 2, \dots, n \end{aligned}$$

Where s^- false is the input slack variable and s^+ false is the output slack variable. λ_j false is the weight of the j th DMU and θ false means the j th banks' technical efficiency value (TE). If $\theta = 1$ false, and $s^- = 0$ false, $s^+ = 0$ false, DMU_{j_0} false is DEA efficient which means those commercial banks has achieved the biggest outcome via their current CSR input; If $\theta < 1$ false, then DMU_{j_0} false is DEA inefficient.

3.2.2 Corpus Tool

This paper mainly adopts the third generation corpus tool---Antconc's keyword function. Keyword analysis is one of the core techniques of corpus linguistics which refers to a word that is more important to a text or a culture than another text or culture (Hu, 2015). Using a reference corpus maybe useful in revealing those words that are under-represented in the observe or the target corpus. Antconc gives a list of all the negative keywords-list which did not occur in the target corpus, but much less frequently than would be expected by chance alone when compared to the reference corpus. (Baker, 2006) For example, if a word oc-

curs comparatively more often in, say, a corpus of modern English children's stories, when compared to the British National corpus, we conclude that such a word has high saliency, keyness in the genre of children's stories and is worth investigating in further detail. According to the purpose of the study, this paper uses the CSR report of CSR efficient commercial bank as an observation corpus, and the CSR report of CSR inefficient commercial bank as a reference corpus, and compares the two to find out the salient or significant word---the keyword used in CSR efficient commercial banks' CSR.

Given that "a keyword is only available to researchers in the form of language, in order to answer specific research questions, researchers must carry out in-depth explanations" (Baker, 2004). Therefore, after the keyword extraction, they should be placed in the context of the CSR report and combined with the concordance function of Antconc to make a qualitative study of quantitative language patterns (Hu, 2015) in order to reveal discourse characteristics of CSR efficient banks' CSR report.

3.3 The selection of indicators of DEA and corpus

3.3.1 The selection of DMUs

Table 1 China Commercial Bank's Total Asset Ranking in 2015

Unit: Million

DMU	Classification of Bank	Total Asset	Rank
Industrial and Commercial Bank(IC-BC)	Large state-owned commercial bank	21,034,098	1
China Construction Bank(CCB)	Large state-owned commercial bank	17,890,733	2
Agricultural Bank of China(ABC)	Large state-owned commercial bank	17,686,237	3
Bank of China(BOC)	Large state-owned commercial bank	14,786,678	4
Bank of Communications(BOCOM)	Large state-owned commercial bank	7,014,471	5
China Merchants Bank(CMB)	National joint-stock commercial bank	5,474,978	6
China Industrial Bank (CIB)	National joint-stock commercial bank	5,185,434	7
Shanghai Pudong Development Bank (SPDB)	National joint-stock commercial bank	4,984,518	8
China CITIC Bank (CITIC Bank)	National joint-stock commercial bank	4,884,295	9
China Minsheng Bank(CMBC)	National joint-stock commercial bank	4,357,468	10
China Everbright Bank(CEB);	National joint-stock commercial bank	3,132,315	11
Ping An Bank(PAB)	National joint-stock commercial bank	2,507,149	12
Huaxia Bank (HB)	National joint-stock commercial bank	1,984,821	13
China Guangfa Bank (CGB)	National joint-stock commercial bank	1,836,587	14
Bank of Beijing(BOB)	City Commercial Bank	1,826,404	15
Bank of Shanghai(BOS)	City Commercial Bank	1,433,190	16
Bank of Jiangsu (JSBC)	City Commercial Bank	1,283,454	17
Heng Feng Bank(HFB)	National joint-stock commercial bank	1,063,244	18
China Bohai Bank(CBB)	National joint-stock commercial bank	1,031,650	19
Bank of Nanjing(NJCB)	National joint-stock commercial bank	80,5020	20

From the data in the figure 1, we can see that large-scale state-owned commercial banks and national joint-stock commercial banks are the two types most representative commercial banks with the largest assets in China's banking industry. Therefore, 5 large state-owned commercial banks including Industrial and Commercial Bank, China Construction Bank, Agricultural Bank of China, Bank of China, Bank of Communications and 10 national joint-stock commercial banks including China Merchants Bank, China CITIC Bank, Shanghai Pudong Development Bank, Industrial Bank, Minsheng Bank, Huaxia Bank, Everbright Bank, Guangdong Development Bank, Ping An Bank and Bohai Bank have been selected as the DMU in DEA model. (Heng Feng Bank failed to meet the needs of this study because they did not compile 2015 corporate social responsibility report).

3.3.2. The selection of input indicators for DEA

According to Maignan & Ralson's study, the contents of CSR include philanthropy the listed companies engaged in, charitable donations, volunteer services, compliance with ethical guidelines, quality management, safety plans, and environmental management (Maignan & Ralson, 2001). Therefore, based on the availability and quantification of indicators, this study combines the research results of Maignan et al. and the characteristics of China's commercial banks to select the input indicators for the DEA model which are donation amount, green credit balance, loan balance of small and micro enterprises (SMEs), and cash paid for employees during the year.

In the selection of DEA model output indicators, this paper takes into account the bank's common sense on profitability and asset size in actual operations, and selects net profit and new balance of loans in the current year as a measure of bank profitability and scale expansion capability respectively as the output indicators. As the following table shows.

Table 2 The input and output factors of each commercial bank in 2015

Unit: ten hundred

DMU	X1 (donation amount)	X2 (green credit bal- ance)	X3 (loan balance of SMEs)	X4 (cash paid for employee)	Y1 (net profit)	Y2 (new balance of loans)
ICBC	5,575	70,280,000	188,320,768	10,393,300	26,326,500	86,481,406
ABC	4,867	54,313,100	108,822,798	10,970,700	18,077,900	82,346,233
BOC	5,900	41,231,500	114,571,373	6,892,700	15,219,900	66,288,806
CCB	4,121	73,356,300	127,787,861	8,882,200	22,517,600	99,897,185
BOCOM	2,448	20,479,500	62,445,786	2,325,700	6,357,400	28,757,061
CITIC Bank	1,559	2,369,600	44,198,945	2,048,100	3,967,200	30,191,172
CEB	613	3,870,000	28,843,523	1,264,000	2,914,400	21,641,906
HB	2,273	3,996,000	23,618,731	1,160,000	1,858,100	12,010,895
CGB	917	341,400	16,964,615	899,144	906,397	7,633,865
PAB	1,035	1,640,200	21,899,106	1,287,100	2,186,500	19,116,994
CMB	3,647	15,650,300	58,980,093	2,531,300	5,801,800	30,350,078
SPDB	1,263	17,178,500	51,046,902	1,899,500	5,012,700	21,783,457
CIB	3,959	39,420,000	56,053,934	1,771,500	4,788,000	23,171,130
CMBC	6,500	1,140,400	57,222,675	2,212,500	4,548,500	23,600,248
CBB	235	1,119,100	5,454,063	319,177	568,851	6,930,110

(The above data comes from the annual reports of banks and Guotai Junan Database)



3.2.3 The Selection of Corpora

The corpus of this research is the CSR report of 15 major commercial banks in China in 2015. The source is commercial banks' official website. A CSR report corpus of China's major commercial banks with a total capacity of 497,926 tokens and 29,922 types was established, and the segmentation software Segmenter was used to segment the Chinese corpus which is convenient for Antconc to retrieve.

3.4 Research Steps

According to the four research questions in this paper, this study includes the following steps:

(1).The DEA model is used to study the CSR input efficiency of commercial banks, and the DEA efficient and non-efficient commercial banks are distinguished.

(2).Use the reports of DEA efficient banks (CSR-A) as the observe corpus and reports of DEA non-efficient banks(CSR-B) as the reference corpus in order to examine the former's keyword list with reference to the latter.

(3).Combining with Antconc's concordance function, the keyword the CSR-A corpora against CSR-B corpora were examined in the semantic context of the CSR report, and the underlying causes of the keyness were analyzed based on the evaluation criteria for sustainable development of commercial banks. Qualitative research was conducted and feature of keyword of DEA efficient banks summarized and analyzed from the text.

(4).According to the results of qualitative research, suggestions are made for China commercial banks to increase the efficiency of CSR input in order to achieve sustainable development.

4 Findings and Discussions

4.1 The DEA Empirical Results of the Effectiveness of CSR input of China Major Commercial Banks

Table 3 Effectiveness of CSR input of China Major Commercial Banks based on CCR Model

DMU	Technical Efficiency (TE)	Rank	DEA Effective- ness
China CITIC Bank	1.000	1	efficient
Huaxia Bank	1.000	1	efficient
Guangdong Development Bank	1.000	1	efficient
China Merchants Bank	1.000	1	efficient
Industrial Bank	1.000	1	efficient
Bohai Bank	1.000	1	efficient
Everbright Bank	0.798	7	inefficient
Minsheng Bank	0.684	8	inefficient
Bank of Communications	0.676	9	inefficient
China Construction Bank	0.417	10	inefficient
Industrial and Commercial Bank	0.298	11	inefficient
Bank of China	0.215	12	inefficient
Ping An Bank	0.202	13	inefficient
Agricultural Bank of China	0.173	14	inefficient
Shanghai Pudong Development Bank	0.173	15	inefficient

Through the results shown in the table above, we can see that only 6 of the 15 commercial banks are DEA efficient. They are China CITIC Bank, Huaxia Bank, Guangdong Development Bank, China Merchants Bank, Industrial Bank and Bohai Bank. The remaining nine commercial banks, including the Big Four bank in China, are DEA inefficient. Among them, Bank of China, Ping An Bank, Agricultural Bank, and Shanghai Pudong Development Bank ranked lower, and there is a large gap from the ideal efficiency value.

4.2 Keyword Retrieval and Analysis of Commercial Bank CSR Reports

Based on the above empirical results, the study used the CSR reports of 6 DEA efficient commercial banks as an observational corpus CSR-A containing 135,936 tokens and 10,947 types; The CSR report of 9 DEA inefficient commercial bank is built as a reference corpus CSR-B containing the 361,990 tokens and 18,948 types is established. The size of the CSR-B is about three times that of the CSR-A, which is in line with the scale of the reference corpora.

Calculate keyness by selecting statistical indicators of Log-likelihood in Antconc (we select a p-value of 0.05 as the cut-off point in determining whether a word was key or not) and rank the keyword according to Keyness from highest to lowest. Some proper nouns which contains “Huaxia”, “Merchant” etc. and other place names like “Fuzhou”, “Shaoxing” are removed and the top thirty keyword are selected. Then we get the following keyword list of DEA efficient banks’ CSR reports relative to DEA inefficient banks’:

Table 4. The Rank of Keyword list CSR-A against CSR-B

Rank	Freq	Keyness(LL)	Keyword
1	734	83.561	我们
2	307	68.465	为本
3	67	56.503	赤道
4	55	53.370	为您
5	2059	51.463	客户
6	57	44.316	于心
7	54	43.896	实现
8	2098	41.357	社会
9	75	39.777	零售
10	46	38.601	责任
11	87	38.458	战略
12	508	35.231	员工
13	122	29.683	互联网
14	25	28.239	轻型
15	169	25.823	小微
16	36	20.402	市场
17	23	18.074	培训
18	318	14.763	理念
19	18	13.403	社区
20	24	12.079	刷脸
21	79	12.032	绿色
22	180	11.420	成长
23	54	10.629	零售
24	17	9.139	追求
25	126	8.539	和谐
26	58	7.577	努力
27	16	7.232	云闪付
28	373	6.731	服务
29	951	6.234	风险
30	20	6.194	老年人

In 2006, in order to recognize the banks that take social, environmental, and corporate governance goals into account in business operations and play a leading and innovative role in these aspects, the “Financial Times” of the UK and the International Finance Corporation (IFC) jointly established the “Sustainable Banking Award”. The purpose of this award can be seen as the elaboration of the bank’s sustainable development connotation. The Industrial Bank of China has won this award twice. Regarding the measurement and evaluation of sustainable development of finance and banking, Zhou designs a set of evaluation index system of banking financial situation based on sustainable development from six dimensions, which are safety, liquidity, profitability, development, environmental responsibility, and social responsibility. And he uses the analytic hierarchy process to determine the weight of each index. In addition, the measurement of bank competitiveness and comprehensive evaluation also have important reference value for the analysis of the factors that influence the sustainable development of commercial banks (Zhou, 2008). The “China Commercial Bank Competitiveness Evaluation Research Group” of “The Banker” magazine (2006) proposes the competitiveness model of China’s commercial banks and defines the competitiveness process as core competence in seven areas, including development strategy, corporate governance, products and services, information technology, risk management, human resources, and market influence. In terms of ways and strategies for commercial banks to achieve sustainable development, He proposes that the driving force for sustainable financial development comes from financial reforms, financial innovations, and financial incentives (He, 2000). Sun & Li believe that the bank’s sustainable development strategy should be implemented by integrating the green concepts and requirements of environmental protection, energy conservation and emission reduction into the bank’s main business operations and investment decisions (Sun & Li, 2011). The former Agricultural Bank of China’s President Xiang believes that, “to achieve the sustainable development of commercial banks is to follow the general trend of financial globalization, liberalization, and comprehensive development, and actively promotes the pace of business transformation from various areas such as the development strategy, business processes, business structure, financial innovation seek to improve their competitiveness”. According to academic research’s connotation, evaluation methods and implementation of sustainable development of commercial banks, the performance of commercial banks in terms of profitability, operation and management, development mode, innovation capability, governance structure, and corporate culture is the determinant factors of the capabilities of the sustainable development.

Based on the above analysis, this paper finds that the keywords of CSR efficient commercial banks versus inefficient commercial banks in Table 4 coincides with key factors determining the sustainable development of commercial banks. To facilitate the following text analysis, the above 30 keywords are divided into the following 6 categories according to their semantic domains and the connotation and principles of sustainable development, as shown in Table 5. Therefore, this paper considers that commercial banks with efficient CSR input have achieved a better sustainable development in terms of corporate culture, stakeholders, marketization level, application of science and technology, innovative business, and external supervision mechanisms, compared to ineffective commercial banks.

Table 5. The keywords list of CSR report of the DEA-efficient commercial banks

Classification	Keyword
Corporate culture	We, root, heartedly, value, achieve, pursue, for you, sincerely, the aged, endeavor, service
Stakeholders	Employee, client, growth, harmony, training, community, society
Marketization level	Market, risk, strategy, light
Technology applications	Internet, face swiping, cloud payment
Innovation business	Green, small and micro, retail
External monitoring mechanism	Equator

4.2.1 Keywords analysis of corporate culture

CSR efficient banks integrates in its corporate culture and service philosophy into its CSR value that rewards stakeholders, which will have a profound impact on banks' sustainable management and development. From table 4 we can see the keyword "we" ranked the first with the keyness of 83.561 in CSR efficient banks' CSR report against the inefficient banks'. "we" and the forth ranked keyword "for you" (keyness:53.370) belongs to the self-mention and engagement markers of interactional metadiscourse. Metadiscourse refers to "language resources used to organize texts and indicate the author's position on the content of the discourse and the reader's position." (Hyland, 2005) The interactional metadiscourse is evaluative and interventional. It directly affects the closeness of the relationship between the author and the reader and the expression of the author's attitude. The self-mention "we", the high frequency of self-mention "we"-first-person pronouns appeared in the CSR report of DEA efficient commercial bank, showing that such banks have narrowed the distance with stakeholders through CSR reports, formed alliances with stakeholders, and viewed the needs of stakeholders as their own needs. High-frequency use of the first-person pronoun "we" is a reflection of the interpersonal meaning in this type of CSR report. Callow believes that the first person indicates that the author is passing on beliefs rather than facts, which will reduce the veracity of the text, but it will help establish a relationship between the company and the reader (Callow, 1998). For example, Bohai Bank's CSR report mentioned: It can be seen that Bohai has actively worked hard to establish a positive, responsible and credible corporate citizenship through the use of the first personal pronoun "we" in the CSR report.

“我们积极投身灾难救助，捐款援助天津“8·12”爆炸事故牺牲、受伤消防战士和受灾群众，发挥金融企业优势支持善后工作；”

It can be seen that Bohai Bank has actively worked hard to establish a positive, responsible and credible corporate citizenship through the use of the first personal pronoun "we" in the CSR report.

As for the forth ranked keyword "for you" (keyness:53.370)-engagement markers of interactional metadiscourse, "您" is the honorable title of "you" in Chinese culture which shows banks' respective and sincere stance and attitude in front of their stakeholders. The words "for you" and "you" and "sincere", "the aged", "effort", and "service" are keyword in the CSR efficient banks' CSR report. It shows that such banks are striving to strengthen their communication with stakeholders through the concept of "customer first" and differentiated services for different ages and levels of stakeholders which lead to banks' improvement of both CSR and operation efficiency and the achievement of sustainable development. As Huaxia bank and Industrial bank says in its CSR report:

“华夏银行秉持‘一心为您’的服务理念，以支持实体经济的健康快速发展作为出发点和落脚点，不断创新金融产品和业务模式，服务国家‘一带一路’战略，探索自贸区服务模式，支持中小微企业成长和中国企业‘走出去’，推动区域协调发展和产业转型升级，实现地方经济社会的可持续发展。”

“兴业银行始终倡导‘寓义于利’的社会责任实践方式，以商业行为承担社会责任。“安愉人生”就是在充分调研老年人需求的基础上，通过商业模式的突破，将履行社会责任与自身业务发展相结合，体现了对老年人的关爱。”

A good corporate culture plays a role in strengthening and reinforcing the concept of social responsibility of commercial banks. The keyword "root", "heartedly" are the reflection of CSR efficient Commercial Bank's business concept and corporate culture ---"Corporate Reward Society". "聚焦于新 履责于行 融责于基 融责于心" are Huaxia Bank's corporate culture advocated in its CSR reports. "以人为本"也是华夏银行始终坚持的核心思想：“我们坚持以人为本，关爱员工发展，健全职工代表大会制度，疏通员工职业发展通道，将银行发展成果多维度惠及员工、股东、客户、社会。” A good corporate culture of the bank will create a sense of "identity" and "a sense of belonging" from the heart of the bank employees which consciously links the individual's future with the bank's future and enables employee develop individual talents actively, and improves service quality. This will effectively promote commercial banks to improve their scientific management, enhance cohesion and creativity so as to achieve sustainable



development. In other words, only when a good cultural atmosphere is established, a commercial bank can actively take various social responsibility obligations in its business activities, and can consciously combine corporate profits with the interests of stakeholders in all aspects of decision-making. (Zhu, 2004)

4.2.2 Keywords analysis about stakeholders

Compared with the DEA-inefficient banks, the DEA-efficient banks in pay more attention to meet the expectations and needs of stakeholders, which can directly and indirectly contribute to the sustainable development of banks. The keyness of “customer” is up to 51.463 ranking the fifth in DEA efficient banks’ CSR report against the inefficient one and the keyness of stakeholder related keywords “society”, “employee” are higher than 30. In addition, keywords such as community, training, and growth are related to commercial banks’ stakeholders, and they all reflect the appreciation of the DEA-efficient banks to the interests of stakeholders. According to the theory of stakeholders (Freeman, 1984), the company’s main stakeholders are the stakeholders who play a fundamental role in the daily operations of the company, including shareholders, creditors, employees, clients, suppliers, etc, and the secondary stakeholders refers to stakeholders that can influence the main stakeholders, including communities and non-profit organizations. Satisfying the needs of key stakeholders can bring direct benefits to the company when profitability is the basis for the sustainable development of commercial banks. However, satisfying the needs of secondary stakeholders by actively fulfilling corporate social responsibilities can indirectly bring sustainable economic profit to commercial banks. Bohai Bank states in its CSR report that, “我们倡导快乐工作, 幸福生活, 促进员工发展, 真挚关爱员工, 打造优秀企业文化, 建设和谐企业”、“本行通过完善薪酬激励制度和福利体系, 以及建立完备的员工职业生涯发展规划, 满足员工获得稳定合理的收入, 享有充分的权益保障, 以及良好的职业生涯规划 and 成长机会的价值要求”和“本行高度重视社会责任管理, 面向股东、客户、员工、合作者、社会等利益相关方的期望和要求, 形成了自己的社会责任观, 明确了责任履行方向和责任实践方向”。Huaxia Bank states in its CSR report that, “我行在构建全面覆盖培训的基础上, 又提出具有针对性的培训体系, 积极培育具有现代管理意识和专业水准的多元化人才队伍”。

4.2.3 Keywords analysis about marketization level

The DEA efficient banks have improved the bank’s operating efficiency and marketization level through scientific and reasonable governance structures, management mechanisms, strategic objectives, and risk management. Commercial bank as a special public enterprise, its corporate governance structure is not only related to the bank’s own business development, but also related to the vital interests of the public, so that commercial banks should implement sustainable development in optimizing the organizational structure and internal control. The keywords such as “market (keyness:20.402), risk(keyness: 6.243), strategy(keyness: 38.458), light(keyness: 28.239), and pursuit (keyness: 9.139) have reflected the higher degree of commercialization of DEA efficient commercial banks. For “strategy”, Industrial Bank states in its CSR report that“立行以来, 兴业银行一贯秉承市场化的经营导向”, which reflects the principle that the DEA efficient banks adhere to market-oriented management. Industrial Bank makes the statement that “进一步完善风险管理组织体系及配套运作机制, 全面加强全行风险内控“三道防线”, 充分发挥风险管理对业务发展的保驾护航作用”, which shows that the DEA efficient banks attach importance to the construction of risk management and internal control systems, which is conducive to improving their marketization level. For the “strategy,” Industrial Bank states in its CSR report that “为国家战略重大项目提供多元化融资支持”, while China Merchants Bank makes the statement that“我们顺应国家战略要求, 支持企业走出去和引进来”, which shows that the DEA-efficient banks can follow the national strategies in their operations and keep up with the hotspot of the market. For “light”, China Merchants Bank states in its CSR report that“确立了“轻型银行”的战略方向和“一体两翼”的战略定位, 努力以更少的资本消耗、更集约的经营方式和更灵巧的应变能力, 有效动员和配置金融资源, 提高金融服务效率”, which reflects the bank’s strategic goal of striving to improve operational efficiency, and it further promotes the improvement of its marketization level. On the whole, the DEA-efficient banks’ operations are more

market-oriented than the DEA-inefficient banks. The DEA-efficient banks have achieved sustainable development through rational resource allocation and efficient business processes.

4.2.4 Keywords analysis about technology applications

In order to seize the competition opportunities to promote sustainable development, the DEA-efficient banks in CSR input increase the use of the Internet in business development, actively promote the “Internet+” development model, and create innovate financial service tools and financial service concepts to cope with the challenges of commercial banks in the new situation. The keyness “Internet, brush face, cloud payment” in DEA efficient banks’ CSR report is 29.683, 12.079, 7.232 respectively, among which “Internet” ranks 15th more than most of the keywords related to basic sustainable development concepts. It shows that the use of Internet technology has become an important means for commercial banks to fulfill corporate social responsibility and achieve sustainable development. For example, Industrial Bank makes the statement that “更深入推进互联网金融、大数据、云服务管理的相关运营管理体系建设, 以及社会金融信息 and 产品流量多渠道对接、线上线下优势互补的相关经营布局, 本行集团数字化、信息化、智能化水平有较大提升”, China Merchants Bank states in its CSR report that “在互联网金融正在颠覆传统金融业的同时, 我们加快推进互联网金融的布局与创新, 以“网上转账全免费”、“刷脸取款”等创新创举, 为客户提供更实惠、更便捷、更人性化的业务功能和体验”, and China Guangfa Bank states in its CSR report that “我们顺应客户生活行为移动化、网络化和碎片化演变趋势, 不断升级网上银行服务功能, 为业务办理提供便捷; 基于 HCE 技术与云端支付技术, 推出“云闪付”服务, 丰富移动金融支付体验”。Through the introduction of advanced internet and mobile payment technologies in daily operations, the DEA efficient banks have effectively reduced operating costs, saved a great deal of manpower and material resources, fulfilled corporate social responsibilities, and increased their efforts in providing more convenient and efficient services to their customers, which helps to achieve sustainable development.

4.2.5 Keywords analysis about innovation business

The DEA efficient banks in CSR input actively innovate the development model of green finance, small and micro credits and inclusive finance, realizing the rational allocation of financial resources in various industries, and promoting the dual sustainable development of commercial banks and the entire society. “Green (keyness: 12.032, SMEs (keyness: 25.823, retail (keyness: 10.625)” those keywords reflect the DEA efficient banks’ preference of business areas involving corporate social responsibility fulfillment and sustainable development. Green credit is an important measure of fulfilling corporate social responsibility through the implementation of credit policies by commercial banks to promote the development of environment-friendly companies. China Guangfa Bank states in its CSR report that “2015 年, 我行专门制定了《2015 年广发银行绿色信贷政策指引》, 对我行的绿色信贷项目总体授信策略、客户准入标准和授信要素等做了详细的规范和划分, 将绿色信贷列为“支持类”行业, 引导全行增加绿色经济信贷投放”, which reflects the strong support for green credit. In the context of the government’s strong support for the development of small and micro enterprises, to make small and micro loans is the most effective way for commercial banks to support the development of small and micro enterprises, and to solve the problems of employment and social development. China Citic Bank states in its CSR report that “2015 年, 为有效改善小微企业金融服务, 在小微贷款风险日益高发和多发的形势下, 本行始终高度重视小企业金融业务的发展, 将小企业金融业务作为普惠金融发展、履行社会责任的重点业务, 明确在有效控制授信风险的前提下, 按照“小微化、标准化、工厂化、网络化”的发展思路, 稳健推动小企业金融业务发展, 持续探索和优化小企业金融服务可持续发展的商业模式”, which reflects the unremitting efforts of the DEA-efficient banks in supporting small and micro enterprises and advancing inclusive finance. The retail business is the main carrier for commercial banks to provide good services to the public, so that their degree of emphasis directly affects commercial banks’ service quality to the public and the sustainable development of commercial banks. Industrial Bank states in its CSR report that “为满足社区不同客户群体的差异化金融服务需求, 社区银行提供了特色鲜明、凸显个性的四大零售业务, 使社区居民有了



贴心的金融理财帮手,更是体现了兴业银行立足社区、“便民、利民、亲民”的服务宗旨和服务理念”, which shows its high attention to the development of retail business. The banks mentioned above have increased their performance in fulfilling their corporate social responsibilities by increasing their inclination to key business areas involving corporate social responsibility.

4.2.6 Keywords analysis about external monitoring mechanism

The subject phrase “Equator” reflects that the DEA-efficient banks in CSR input have accepted more advanced external regulatory mechanisms to promote the sustainable development of commercial banks. The Equator Principles are a set of unofficial and voluntary principles developed by the world’s leading financial institutions in accordance with IFC’s environmental and social policies and guidelines. They are designed to identify, assess and manage the environmental and social risks involved in the project financing process. Industrial Bank states in its CSR report that “作为我国首家赤道原则银行, 本行积极贯彻国家低碳经济政策, 将绿色理念贯穿于运营管理的方方面面”, which shows that DEA-efficient banks can improve corporate social responsibility performance through the introduction of international advanced external regulatory mechanisms.

In summary, through the analysis of the keywords of the CSR report of the DEA efficient against DEA inefficient banks’, we have found that the DEA efficient banks achieve better sustainable development than DEA inefficient banks in the shaping of corporate culture, optimization of service attitudes, full-service for stakeholders, improvement of marketization level, acceleration of technology applications, development of innovative businesses, and acceptance of external monitoring mechanisms.

5 Suggestions and Implications

Under the background of the “new normal” economic shift, all industries’ growth mode and development momentum need to be transformed, and all industries have to increase total factor productivity to support the efficiency and sustainable development of China’s economy. As the most important financial intermediary in China, the improvement of commercial banks’ management efficiency and CSR input efficiency will not only benefit the sustainable development of commercial banks themselves, but also optimize the supply of capital for both the the element end and the production end of other industries, which in turn promotes sustainable development of the entire society.

By applying the efficiency DEA model and corpus analysis tool Antconc into the analysis of Corporate Social Responsibility report (CSR report) and taking commercial banks’ performance of CSR as an input which may bring banks effective output, this research investigates CSR input-output efficiency and the achievement of sustainable development of China’s major commercial banks by means of qualitative, quantitative and comparative analysis on the basis of the outcome of DEA empirical study. DEA model finds that the CSR input efficiency of 6 smaller commercial banks including China Citic Bank (CITIC Bank), Huaxia Bank (HB), China Guangfa Bank (CGB), China Merchants Bank (CMB), China Industrial Bank (CIB) and China Bohai Bank (CBB); the other 9 commercial banks including the 4 big banks in China are the inefficient banks. Then 30 keywords and 6 semantic domains of DEA efficient banks’ CSR reports are summarized relative to DEA inefficient banks’ through corpus tool’s keyword analysis. Based on discursive context of CSR reports and characteristics of commercial banks sustainable development, this research finds that DEA efficient banks achieve a better sustainable development than DEA inefficient banks in corporate culture, stakeholder concern, marketization level, application of Fintech, green finance and business innovation reference to exterior supervision mechanism. Finally, based on the research results, combined with the requirements of the supply-side reform under the new normal economy, this article proposes the following suggestions to improve the CSR input-output efficiency of China’s commercial bank, so that commercial banks can better achieve sustainable development:

First, commercial banks should improve service quality and attitude by implementing the service con-

cept of “customer first”, “employee-centered”, and “giving back to society” in the corporate culture. Banks should provide differentiated and innovative financial services for different stakeholders. Second, commercial banks should speed up the process of the marketization, reduce the proportion of national ownership, diversify the ownership structure, and formulate scientific and rational development strategies to achieve maximum utilization efficiency of CSR input. Third, speed up the integration of the Internet and finance. Commercial banks should introduce Internet technologies in financial services and product innovation to reduce costs and increase efficiency. Fourth, the business scope of commercial banks should be tilted toward inclusive finance, green industries, science and technology innovation industries, and small and micro enterprises. Fifth, China’s commercial banks should take the initiative to learn from foreign banks that perform social responsibilities well, actively join the international social responsibility organization, adopt the “Equator Principles”, and establish an effective external supervision mechanism to improve the level of fulfillment of their social responsibilities.

Although the combination of economic model and corpus technology has added objectivity and scientificity to the text analysis and conclusion of this study, this study also has some limitations and deficiencies. In terms of the selection of research objects, this paper selects according to the level of asset rankings and does not cover some city commercial banks and rural commercial banks that are also actively fulfilling their social responsibilities. In addition, in the selection of input indicators, this study only selected four representative, easily accessible and quantifiable input indicators and did not cover the remaining indicators, which may bias the effectiveness of the DEA efficiency results.

References

- 卞继红, 蔡月祥, 张雷 .2015. 视频制造业上市公司企业社会责任投资效率实证研究 [J]. 财务与金融, (1): 80.
- 丁一兵, 付林 . 2015. 中美大型企业社会责任对其企业效率的影响机制研究 — 基于 DEA-Tobit 两分法的分析 [J]. (6): 21-22.
- 郭丽, 2009. 商业银行的企业社会责任与可持续发展战略关系探析 [J]. 南方金融, (08): 67-68.
- 何五星 . 2000. 金融可持续发展的动力问题研究 [J]. 城市金融论坛, (9): 6-10.
- 胡春雨 . 2015. 主题词分析及其在商务话语研究中的应用 [J]. 广东外语外贸大学学报, (4): 48-52.
- 蔺玉 . 企业社会责任投入的有效性研究 - 基于财务收益的实证分析 . 成都理工大学 [D]:50-60
- 倪中新, 王大中, 武凯文 . 2014. 企业社会责任、所有权属性与企业经营效率 —— 基于《企业社会责任蓝皮书 (2009-2013)》的经验证据 [C]. 第九届 (2014) 中国管理学年会 —— 会计与财务分会场论文集 .
- 苏蕊蕊, 仲伟周, 刘尚鑫 .2010. 企业社会责任与企业效率关联性分析 [J]. 山西财经大学学报, (11): 75-85.
- 宋效中, 马欣, 王志敏 .2013. 我国煤炭上市公司社会责任的效率研究 [J]. 会计之友, (3): 62-64.
- 孙琼 . 2009. 绿色金融视角下的我国商业银行可持续发展研究 . 湖南工业大学 [D]:10-12.
- 孙轶颖, 李琳 . 2011. 绿色信贷与银行可持续发展 [J]. 中国金融, (10): 39-40.
- 周怡 . 基于可持续发展的商业银行财务评价体系研究 [D]. 硕士学位论文, 西南财经大学, 2008.
- 钟宏武, 魏紫川, 张蕙等 . 2015. 中国企业社会责任报告 2014: 以报告促管理 [M]. 北京: 经济管理出版社 .
- 朱文忠 . 2004. 论构建商业银行社会责任价值观的系统工程模型 [J]. 经济界, 50-56.
- 朱文忠 . 2008. 商业银行企业社会责任标准与机制研究 . 中山大学 [D]:24-28.
- Baker P. 2004. Querying Keywords[J]. Journal of English Linguistics, (4) : 346-359.
- Banerjee, S. B. (2008). Corporate social responsibility: The good, the bad and the ugly. Critical Sociology, 34(1), 51-79.
- Bhatia A. 2012. The Corporate Social Responsibility Report: The hybridization of a “confused” genre (2007 – 2011) [J]. IEEE Transactions on Professional Communication, (3): 221-238.



- Callow, K. ,1998. *Man and Message: A Guide to Meaning-based Text Analysis*[M]. Boston: University Press of America.
- Carroll, A. B. 2002. The Pyramid of Corporate Social Responsibility: toward the Moral Management of Organizational Stakeholder[J]. *Business Horizons*, (4): 34.
- Charnes, A., Cooper, W.W. & Rhodes, E. 1978. Measuring Efficiency of Decision Making Units[J]. *European Journal of Operational Research*, (1):429-444.
- Clarkson, P.M., Li, Y., Richardson, G.D., Vasvari, F.P., 2008. Revisiting the relation between environmental performance and environmental disclosure: an empirical analysis. *Acc. Organ. Soc.* 33 (4–5), 303–327.
- Dahlsrud, Alexander. 2008. How Social Responsibility is Defined: an Analysis of 37 Definitions[J]. *Corporate Social Responsibility & Environmental Management*, (1):1-13.
- Fairclough, N. 1989. *Language and Power*[M]. London: Pearson.
- Fuoli, Matteo. 2013. Assessing Social Responsibility: A Quantitative Analysis of Appraisal in BP's and IKEA's Social Reports[J]. *Discourse and Communication*, (1):55-81.
- Freeman R. 1984. *Strategic Management: A Stakeholder Approach*[M]. Boston: Pitman Publishing Inc.
- Huang, X., Zhu, W.Z. Chinese corporations' conception of sustainable development: an innovative view of corpus analysis[J]. *Chinese Management Studies (SSCI)*, Vol. 11 Issue: 1, 2017.
- Hyland, K. 2005. *Metadiscourse: Exploring Interaction in Writing*[M]. London: Continuum.
- Iyigun, N, O. (2013) Corporate Social Responsibility and Ethnics in Management in the light of Sustainable Development. *Handbook of Research on Developing Sustainable Value in Economics, Finance, and Marketing*. U. Akkucuk (Ed.), Hershey PA, USA: IGI Global.
- Luo, L., Tang, Q., 2014. Does voluntary carbon disclosure reflect underlying carbon performance? *J. Contemp. Account. Econ.* 10 (3), 191–205.
- Luke, T. W. (2013). Corporate Social Responsibility: An Uneasy Merger of Sustainability and Development. *Sustainable Development*. 21, 83-91.
- Porter, M. E., & Kramer, M. R. 2006. Strategy & Society -the Link Between Competitive Advantage & Corporate Social Responsibility[J]. *Harvard Business Review*, (12): 78-92.
- Porter, M. (2003). CSR-A Religion with too Many Priests?. Keynote Speech on the Role of Corporate Philanthropy, The European Academy of Business in Society's 2nd Colloquium, European Business Forum. 15, Denmark: Copenhagen Business School.
- Margolis J.D, Walsh J.P. 2001. People and Profits? The Search for a Link between a Company's Social and Financial Performance[M]. Mahwah: Lawrence Erlbaum.
- UN Brundtland Commission Report (1987). *Our Common Future*. Oxford: Oxford University Press.
- Wheeler, D.; Colbert, B. & Freeman, R. E. (2003). Focusing on Value: Reconciling Corporate Social Responsibility, Sustainability and a Stakeholder Approach in a Network World. *Journal of General Management*. 28(3), 1–28.
- Seelos, C. & Mair, J. (2004). Social Entrepreneurship: The Contribution of Individual Entrepreneurs to Sustainable Development. Working Paper. 553, Barcelona: University of Navarra Anselmo Rubiralta Center for Globalization and Strategy Center for Business in Society.
- Mautner, G. (2005), "Time to get wired: using web-based corpora in critical discourse analysis", *Discourse & Society*, Vol. 16 No. 46, pp. 809-828.
- Neu, D., Warsame, H., Pedwell, K., 1998. Managing public impressions: environmental disclosures in annual reports. *Acc. Organ. Soc.* 23 (3), 265–282.
- Wiseman, J., 1982. An evaluation of environmental disclosures made in corporate annual reports. *Acc. Organ. Soc.* 7 (1), 53–63.

Exploring Academic Discursive Competence in Chinese EFL Undergraduates Abstract Writing of Research Articles

Fengjie Zhou¹, Fei Li^{2*}

¹Huazhong University of Science and Technology, zhoufengjie2018@163.com

²China University of Geosciences (Wuhan), fei.li@cug.edu.cn

*Corresponding author Email: fei.li@cug.edu.cn

Abstract

Effective abstract writing is critical for Chinese EFL undergraduates to succeed in academic publishing and global scholarly communication, yet their academic discursive competence in this area remains underexplored. Despite growing research on Chinese EFL learners' academic writing (e.g., Li, 2020), few studies focus on undergraduate English majors' research article abstracts through the lens of academic discursive competence. This study addresses this gap by applying Jiang Feng's (2021) academic discursive competence framework and grounded theory methodology to analyze English abstracts written by 80 senior undergraduate English majors at a university in central China. Through open, axial, and selective coding, the analysis examines generic, textual, and social competence. Findings reveal significant variations in abstract move structures, with four unconventional models identified (e.g., omitting research method descriptions). Textually, students favor simple and complex sentences, present tense, and active voice, but exhibit frequent grammatical errors and redundant descriptions. Socially, disciplinary identity markers (e.g., field-specific terminology) are underused, while proximity expressions (e.g., first-person pronouns) prevail over positioning expressions (e.g., objective statements). Using grounded theory, this study identifies context-specific features of academic discursive competence, offering new insights into Chinese EFL learners' abstract writing challenges. These findings provide practical implications for enhancing teaching strategies, improving grammatical accuracy, and fostering disciplinary identity in undergraduate English programs.

Keywords: Academic discursive competence, Research article abstract writing, Chinese EFL undergraduate learners, Move structure analysis



1 Introduction

Academic English writing is vital for internationalized foreign language education, enabling learners to participate in global academic and professional discourse (Hyland, 2005; Swales, 1990). Metadiscourse, defined as linguistic strategies that organize texts, signal structure, and engage readers, is essential for crafting coherent and persuasive academic texts (Hyland & Tse, 2004). Chinese EFL undergraduates face unique challenges in effective academic writing due to differences in the Chinese and English way of thinking and writing (Bian & Wang, 2016). These challenges are evident in research article abstracts, a genre requiring precise metadiscourse to balance conciseness and audience engagement. This study investigates Chinese EFL undergraduate English majors' metadiscourse competence in abstract writing, a critical skill for global academic participation.

Despite the importance of metadiscourse for textual coherence and reader engagement, research on Chinese EFL undergraduates' metadiscourse competence remains limited. Existing studies (e.g., Ren & Li, 2011; Zubir et al., 2021) focus primarily on postgraduates' metadiscourse in genres like dissertations or journal articles, overlooking undergraduates who face distinct challenges due to limited genre exposure and early-stage academic training (Geng et al., 2004). For instance, undergraduates often struggle with using hedges (e.g., “possibly,” “might”) and boosters (e.g., “clearly,” “definitely”) to signal structure or assert arguments in abstracts, reflecting lower genre awareness. This research gap hinders the development of evidence-based pedagogical strategies to enhance their global academic competitiveness (Sun, 2017).

To address this, this study investigates Chinese EFL undergraduate English majors' metadiscourse competence in research article abstract writing. By integrating Hyland's (2005) metadiscourse model with Jiang's (2021) academic discursive competence framework, it seeks to uncover how these students employ hedges and boosters to construct persuasive academic discourse. The findings are expected to inform evidence-based pedagogical strategies, enhancing Chinese EFL undergraduates' academic writing skills and contributing to more effective teaching practices in foreign language education.

2 Literature Review

2.1 Theory of Academic Discursive Competence

Academic discursive competence, rooted in Hymes' (1972) communicative competence, emphasizes the appropriate use of language in social and disciplinary contexts. Bhatia (2004) defines “discursive competence” as comprising three dimensions: textual competence (mastery of linguistic codes and coherent text production), generic competence (constructing genre-specific structures), and social competence (using language to engage in professional activities and build disciplinary identities). Jiang (2021) adapts this framework for academic writing, highlighting its role in fostering international academic exchange and disciplinary identity construction, defined as aligning language use with the norms and values of a specific academic community (Hyland, 2012). This study targets novice Chinese EFL learners, whose limited exposure to academic norms poses unique challenges, unlike Jiang's (2021) focus on experienced writers.

To ensure methodological rigor, this study employs Charmaz's (2006) constructivist grounded theory to analyze academic discursive competence in Chinese EFL contexts. The iterative process—open coding (identifying linguistic patterns, e.g., hedges like “may”), axial coding (linking patterns to genre structures, e.g., rhetorical moves), and selective coding (integrating findings into a refined framework)—enables data-driven insights from student abstracts. Coding reliability is ensured through a team of two coders independently coding 20% of the data, achieving 85% inter-coder agreement (Cohen's kappa), with discrepancies resolved through discussion. Theoretical saturation, as per Corbin and Strauss (2015), is reached when no new categories emerge, ensuring robustness. To address grounded theory's subjectivity, the study employs coder training and triangulation with literature (e.g., Hyland, 2005; Canagarajah, 2002).

This approach refines Jiang's (2021) framework by incorporating cross-cultural genre theorization, as called for by Bhatia (2004). For Chinese EFL learners, cultural influences, such as Confucian values emphasizing respect for authority, may lead to overuse of assertive language, complicating adherence to academic norms of modesty (Wang, 2013). While Bhatia's framework is comprehensive, it lacks specific guidance for non-native English learners (Canagarajah, 2002). This study addresses this gap by focusing on novice Chinese EFL learners, using grounded theory to iteratively refine Jiang's framework for cultural relevance. The following section examines research article abstracts as a key site for applying this framework in Chinese EFL contexts.

2.2 Research Article Abstracts

Research article abstracts are concise summaries that enable readers to assess content relevance and engage with disciplinary discourse (Salager-Meyer, 1990). As structured texts, abstracts provide an ideal context for novice EFL learners to practice and develop academic discursive competence (Rowley, 1988). Recent studies have explored linguistic features (e.g., hedges, nominalization) and cross-disciplinary variations in abstracts (Hyland & Milton, 1997; Kurniawan et al., 2019; Liu et al., 2024; Huang et al., 2025). Hyland and Milton (1997) highlight challenges in EFL writing, noting that cultural preferences for indirectness or formulaic expressions among Chinese learners may hinder alignment with academic norms. Kurniawan et al. (2019) analyze linguistic features across disciplines but use small, discipline-specific samples, limiting generalizability to EFL contexts like China.

Liu et al. show that there is an obvious difference in the integrity of the moves between examine translators' strategic deployment of stance devices across moves in RAAs, focusing on the interplay between micro-level lexico-grammatical choices and macro-level discursive outcomes, thereby revealing translators' professional agency in academic translation.

In recent years, experts and scholars at home and abroad have been expanding the fields and dimensions of abstract research, ranging from hedges, stance adverbs, verb morphological features, linguistic complexity, nominalization, English translation of unattributed sentences, and to comparative studies of abstracts between Chinese and foreign dissertations, between domestic core journals and master's dissertations, and between different disciplines, to name but a few.

However, most of the previous literature is limited to the analysis of the discourse and genre of abstract writing. As the study of academic abstracts deepens, scholars gradually shift the focus from the discursive structure and linguistic features of abstracts to the comprehensive academic discursive ability embodied in abstracts. Yet defining what type of discursive competence is essential for efficient academic communication and learning remains difficult. Jiang (2021) focused on the abstract of the research article in his study, citing the theory of discursive competence proposed by Bhatia (2004). He develops it into corresponding three attributes of academic discursive competence, namely textual competence, generic competence, and social competence.

Similarly, Cai (2022) designed a framework based on genre analysis, syntactic and lexical analysis, meta-discourse analysis, and a set of explicitly scaffolded teaching methods. This teaching method basically fits with the three levels of academic discursive competence proposed by Jiang (2021). Under such cultivation, his participants improved their overall writing ability of structure, their ability to consciously use linguistic resources to express their positions and opinions, their ability to comply with academic norms and use citation strategies, and their critical thinking ability to identify problems in a small amount of time. The above empirical studies illustrate that the discursive competence of abstracts has a positive effect on the output of texts that fit the disciplinary paradigm, the genre competence of layout, and the construction of disciplinary identity, each of which is crucial for academic writing. The absence of abstract's discursive competence makes it difficult for research articles to "express academic arguments and establish the credibility of ideas within their respective knowledge constructs" (Jiang, 2021).



2.3 Summary

A systematic literature search in Web of Science and Scopus (2015–2024) using keywords like “genre-based writing,” “EFL,” and “abstracts” identified only a few studies on genre-based academic writing in Chinese EFL settings, with just less than 10 focusing on abstracts (e.g. Jiang, 2021; Cai, 2022; Zhao, 2023).

From the above findings, it is clear that although the scope of abstract research is wide, there is little research on genre-based abstract writing in domestic academic writing instruction. Cai (2019, 2022) and Jiang (2021) both pointed out the necessity of improving students’ academic discursive competence in abstract writing in their studies. The academic discursive competence of abstracts, however, is still in urgent need of research, and the sorting and construction of core elements of academic discursive competence are still insufficient. This scarcity highlights a critical research gap, particularly given Chinese EFL learners’ challenges with cultural influences (e.g., preference for indirectness) and limited exposure to disciplinary norms (Mauranen, 2012). Therefore, this research conducts an exploratory investigation of the academic discursive competence of Chinese English learners’ abstract writing, which aims to provide a reference resource for sorting out the elements of academic discursive competence, investigating students’ abstract writing competence, and promoting the teaching of academic English writing.

This study fills these gaps by identifying textual (e.g., linguistic accuracy), generic (e.g., move structures), and social (e.g., disciplinary identity) dimensions in Chinese EFL abstracts. However, its focus on Chinese EFL undergraduates may limit generalizability. Future studies could compare Chinese abstracts with those from other EFL contexts, such as Japanese or Korean learners, to identify shared and unique challenges.

3 Theoretical Framework

This study adopts Jiang’s (2021) academic discursive competence framework over Hyland’s (2005) meta-discourse model, as it integrates textual, generic, and social dimensions, aligning with grounded theory’s data-driven approach. Hyland’s (2005) model focuses primarily on metadiscourse, lacking the broader integration of generic and social dimensions critical for analyzing novice EFL learners’ challenges. Grounded theory’s iterative process—open, axial, and selective coding—guides the analysis of 80 Chinese EFL student abstracts. For example, open coding revealed frequent misuse of hedges (e.g., “may” used inappropriately), prompting refinements to Jiang’s social competence dimension to emphasize cultural influences. Theoretical saturation, as per Corbin and Strauss (2014), ensures robustness, while inter-coder agreement (85%) and triangulation with literature (e.g., Hyland, 2005; Canagarajah, 2013) enhance reliability. These insights will inform targeted pedagogical interventions for Chinese EFL learners.

Generic competence involves constructing genre-specific structures, such as rhetorical moves in abstracts (Swales, 1990). Grounded theory coding of 80 Chinese EFL abstracts revealed frequent omission of the “results” move, disrupting genre expectations. Cultural preferences for indirectness may contribute to these gaps (Jiang, 2021), informing pedagogical strategies to teach move structures.

Textual competence focuses on linguistic precision and discipline-specific expression. Coding reveals challenges in vocabulary and sentence structure, such as substituting “prove” for “suggest,” reflecting limited linguistic precision (Jiang, 2021). Cultural preferences for formulaic expressions further complicate textual accuracy, guiding targeted interventions.

Social competence involves constructing disciplinary identity through language (Hyland, 2012). Grounded theory analysis highlights how metadiscourse (e.g., overuse of boosters like “definitely”) reflects Chinese EFL students’ challenges in adopting the modest tone expected in disciplinary communities, addressing limitations in Jiang’s (2021) framework. Cultural influences, such as Confucian respect for authority, may encourage assertive language, necessitating culturally sensitive pedagogy.

4 RESEARCH METHODOLOGY

4.1 Research Questions

Based on Jiang's theory of academic discursive competence (2021), this research adopts a mixed research method. Combined with quantitative and qualitative approaches, this research investigates the current situation of academic discursive competence among senior undergraduate English majors, using the writing of the research article abstract as the entry point. This study aims to answer the following three questions:

- (1) In terms of generic competence, what are the characteristics of moves and steps of Chinese EFL learners' abstract writing of the research article?
- (2) In terms of textual competence, a) what are the linguistic features of Chinese EFL learners' abstract writing of the research article, and b) is the language used standardized?
- (3) In terms of social competence, what are the discursive markers of disciplinary identity used in Chinese EFL learners' abstract writing of the research article?

4.2 Participants

The target participants for this study are senior undergraduate English majors, as they typically have more exposure to academic English writing and have acquired a deeper understanding of academic writing constructs. All participants have passed the TEM-4 exam, with 70% achieving a "good" or "excellent" score. In addition to their general academic background, these students have received some formal training in academic writing, including coursework in writing skills and research methodology.

The study involved 80 students, consisting of 65 females and 15 males. Although the study does not directly analyze gender differences in academic writing performance, it is acknowledged that gender may influence writing styles and content. However, due to the small sample size and the focus on writing competence, this paper does not further investigate gender-related discrepancies in writing performance.

4.3 Data Collection

Data for this study were collected through an abstract writing test. The test followed a four-step process:

- (1) Select the research article and design the background questionnaire. The article chosen was an empirical study on critical thinking and teaching by student teachers, published in an SSCI journal. The length of the article was reduced from 3,620 words to 1,600 words to align with the reading and writing requirements of the Syllabus for the Test for English Majors-Band 8. The reduction mainly focused on the methodology, results, and discussion sections, ensuring that the key information was retained and that students' comprehension and writing abilities were accurately assessed.
- (2) Pilot tests were conducted with two small groups of participants to refine the test procedure and assess the clarity of the article.
- (3) A formal test was administered where students independently read the article and wrote an abstract within 45 minutes, without using external tools.
- (4) Data were collected, organized, and entered into Microsoft Word and Excel for analysis.

4.4 Data Analysis

4.4.1 Analysis of generic competence

Hyland's (2000) move model is chosen for analyzing generic competence due to several key reasons. Firstly, this model is widely recognized and extensively used in academic research across various disciplines, which makes it a reliable and representative framework. Secondly, in comparison with other models, such as Swales's CARS model and Santos's five-move model, Hyland's approach offers a more detailed and



structured analysis with clear subdivisions of steps within each move. This granularity allows for a more precise and nuanced understanding of the abstract writing process. Lastly, Hyland's model, based on a large corpus of 800 abstracts across eight disciplines, ensures its applicability and generalizability across a wide range of academic contexts.

In addition to Hyland's model, Kanoksilapatham's (2005) model, which focuses on the frequency of move usage in abstract writing, is utilized to evaluate the relative importance of each move in the students' writing. This model categorizes moves as compulsory, regular, or optional, depending on their frequency of use, offering a quantitative perspective on the prevalence and relevance of each move.

The analysis of generic competence follows a structured four-step process: entry, separation, tagging, and collection.

(1)Entry: All abstract texts are organized and entered into a digital format for analysis, utilizing software like Microsoft Word and Excel.

(2)Separation: Each abstract is deconstructed into individual sentence units to facilitate detailed analysis.

(3)Tagging: To determine the most appropriate model for this research, a review of previous studies on mainstream move models is summarized in Table 1. This includes the PMRC model created by Graetz (1985), Swales's CARS and IMRD models (1990), Bhatia's IMRC model (1993), Santos's five-move model (1996), and Hyland's refined five-move model (2000). Hyland's (2000) move model is selected for its compatibility, and sentences are tagged according to the relevant moves and steps in Table 2.

(4)Collection: The results of the tagging are aggregated, and each abstract is analyzed to identify the sequence of moves used, as illustrated in Table 3.

Table 1: The Comparison of move models

Move models	Scholar (Year)	Data capacity	Fields	Moves
PMRC model	Graetz (1985)	87	Health Science; Social Science; Education; Humanities	Problems; Methods; Results; Conclusion
CARS model	Swales (1990)		Physics; Biology; Medicine; Social Sciences	Establishing a territory; Establishing a niche; Occupying the niche
IMRD model	Swales (1990, 2001)			Introduction; Method; Results; Discussion
IMRC model	Bhatia (1993)			Introduction; Method; Result; Conclusions
Initial Five-move model	Santos (1996)	94	Applied Linguistics	Situating the research; Presenting the research; Describing the methodology; Summarizing the results; Discussing the research

Upgraded Five-move model (IPMPPrC model)	Hyland (2000)	800	Eight principles in science and social sciences	Introduction; Purpose; Methods; Product; Conclusion
---	---------------	-----	---	---

Table 2: Hyland's (2000) move model

Move	Step	Label
Introduction (M1)	1	Arguing for topic significance or prominences
	2	Making topic generalizations: what is currently known
	3	Defining the key term(s)
	4	Identifying gap
Purpose (M2)	5	Stating general and/or specific purpose of the research including the hypothesis
Method (M3)	6	Describing participants
	7	Describing instrument(s)
	8	Describing procedure and context
Product (M4)	9	Describing the main specific findings of the research
Conclusion (M5)	10	Deducing conclusions from results by commenting on or interpreting the result or deducing claims from the results
	11	Evaluating the significance or contribution of the research
	12	Stating limitation
	13	Presenting recommendation and implication

Table 3: Move-step marking (Research Article abstract no. 17)

Content	Move	Step
Critical thinking is required for students considering its importance nowadays.	M1	S1
Hopefully, the study can shed light on how to develop students' critical thinking.	M5	S11

4.4.2 Analysis of textual competence

Textual competence is assessed by examining linguistic features such as sentence patterns, tenses, and voices. The specific analytical angles are adjusted based on the actual content of Chinese EFL learners' abstract texts, as outlined in Table 4.

Table 4: The classification table of linguistic features in this research

Sentence pattern	Tense	Voice
Simple	Simple present	Active voice
Complex	Simple past	Passive voice
Compound	Simple future	
Compound-Complex	Present perfect	



4.4.3 Analysis of social competence

The analysis of social competence draws on Hyland's (2012) framework for discipline identity markers. These markers are categorized as either "Positioning" or "Proximity" and are detailed in Table 5. The markers in the "Positioning" category express the author's stance on the research, while the "Proximity" markers reflect the relationship between the text and the audience.

Table 5: Markers of discipline identity of "Positioning" and "Proximity"

Category	Function	Example
Positioning		
Hedges	Withhold commitment and open dialogue	Might, perhaps, possible
Boosters	Emphasize certainty or close dialogue	In fact, definitely, must
Attitude markers	Express writer's attitude to proposition	Unfortunately, I agree, surprising
Self mentions	Explicit reference to authors	I, we, my, me, our
Proximity		
Transitions	Express relations between main clauses	In addition, but, thus, and
Frame markers	Refer to discourse acts sequences or stages	Finally, to conclude
Endophorics	Refer to information in other parts of the texts	Noted above, see Fig
Evidentials	Refer to information from other texts	According to X, Z states
Code glosses	Elaborate propositional meanings	Namely, e.g., in other words

4.4.4 Intercoder reliability

The reliability of the coding process is ensured through two rounds of proofreading and annotation. The primary researcher manually identifies the moves and steps, and calculates the frequency and proportion of each data point. To minimize subjective bias, all annotators are trained according to the classification criteria. The consistency between the first and second rounds of labeling is measured at 92.34%, indicating a high level of accuracy and reliability in the analysis process.

5 Results and Discussion

5.1 Analysis of Generic Competence

5.1.1 Occurrence and percentage of moves and steps

This section primarily describes the distribution, occurrence, and percentage of moves and steps. The original author's move and step sequences are as follows: (M1-S4), (M3-S6, S7, S8), (M4-S9, repeated twice), and (M5-S13). The distribution of moves and steps for Chinese EFL learners is presented below.

5.1.1.1 Occurrence of move

The data in Table 6 show that Chinese EFL learners' abstracts do not include any compulsory moves, with their occurrence rate being less than 100%. M1, M2, M3, M4, and M5 are all regular moves, with no optional moves. The most frequent occurrence in Chinese EFL learners' abstracts is M2 (stating the general and/or specific purpose of the study, including the hypothesis), which indicates that Chinese EFL learners are aware of the importance of summarizing their research purpose. The least frequent occurrence in Chinese EFL learners' abstracts is M4 (describing the main specific findings of the study), suggesting that Chinese EFL learners' awareness of describing the results of their research could be improved.

Table 6: The Number and percentage of moves in the abstracts of Chinese EFL learners

Move	M1(Introduction)	M2(Purpose)	M3(Method)	M4(Result)	M5(Conclusion)
N	62	66	64	55	63
P	77.5%	82.5%	80%	68.75%	78.75%

(N = number; P=percentage)

The data in Table 7 show that the largest proportion of Chinese EFL learners' abstract (40%) covers four moves, which is the same as the number of moves of the author's abstract. This shows a tendency to omit moves (omitted move will be discussed further in the unconventional move structure below). In addition, abstracts covering five or three moves are also common, reaching 27.5% and 23.75%, respectively. The data on the percentage of abstracts covering five moves indicate that a significant proportion of Chinese EFL learners are aware of the five-move model of abstract in their writing.

Table 7: The Number and percentage of moves used in Chinese EFL learners' abstracts

Number of moves	Number	Percentage
Five moves	22	27.5%
Four moves	32	40%
Three moves	19	23.75%
Two moves	7	8.75%
One move	0	0

Note: (Repeated moves are not counted)

Compared with the author's abstract, a notable characteristic of Chinese EFL learners' abstracts is that the distribution of the moves is less structured. In the author's abstract, each sentence corresponds accurately to one move, with detailed and concise information focusing on the research methods and results. In contrast, 10% of the Chinese EFL learners excessively focus on M1 (introduction to the topic), with repeated information throughout, reflecting their awareness of the moves but lacking writing skills.

5.1.1.2 Occurrence of step

The step markers show that all steps in the classification Hyland are covered in the Chinese EFL learners' abstracts, while the only step not covered is S12 (stating limitation). Table 8 displays the actual number and percentage of all steps.

Table 8: The Number and percentage of steps in the abstracts of Chinese EFL learners

Step	Number	Percentage
S1	60	75%
S2	5	6.25%
S3	5	6.25%
S4	20	25%
S5	66	82.5%
S6	27	33.75%
S7	48	60%
S8	30	37.5%
S9	55	68.75%
S10	19	23.75%
S11	10	12.5%
S12	0	0
S13	35	43.75%

Among them, Chinese EFL learners' abstracts do not have any compulsory steps, but have regular steps such as S1 (arguing for topic significance or prominences), S5 (stating general and/or specific purpose of the research including the hypothesis), and S9 (describing the main specific findings of the research), indicating that Chinese EFL learners mostly focus on arguing for the importance of the topic of the study, describing the purpose of the study and the main findings. There are optional steps such as S2 (making topic generalizations: what is currently known), S3 (defining the key term), S4 (identifying research gaps), S6 (describing participants), S7 (describing instruments), S8 (describing procedures and context), S10 (deducing conclusions from results by commenting on or interpreting the result or deducing claims from the results), S11 (evaluating the significance or contribution of the study), and S13 (presenting recommendation and implication), suggesting that Chinese EFL learners could also consider the topic definition and research gap identification in the introduction, could introduce participants, research tools and research procedures when describing research methods, and could focus on presenting conclusions and references and making recommendations when analyzing the significance of the research.

There might be two reasons why Chinese EFL learners did not consider the S12: first, they might not be familiar with the method of stating limitations in the abstract, so they did not consider this content. Second, the test paper itself has few or no direct references to the limitations of the content, so they did not address this information when writing the abstract.

5.1.2 Move Sequence

Although the Chinese EFL learners' abstract cover all the moves in Hyland's classification, they show a wide range of variation in move sequences. Therefore, this section looks at the move sequences and the unconventional move structures within them to explore the intention behind the move sequences. Move 1 (M1) is replaced by the number "1" and the move sequence is coded as "1-2-3-4-5", and the move sequence statistics are shown in Table 9.

Table 9: The Number and percentage of step in the abstracts of Chinese

Occurrence of move sequence							
Move sequence	1-2	1-2-3	1-2-3-4	1-2-3-4-1	1-2-3-4-5	1-2-3-5	1-2-3-5-3
Frequency	3	2	3	1	19	6	1
Move sequence	1-2-4-5	1-2-5	1-3	1-3-2-4	1-3-4	1-3-4-5	1-3-4-5-2
Frequency	2	5	1	1	2	7	1
Move sequence	1-3-5	1-4-5	1-5	2-1-2	2-1-3-4	2-1-3-4-5	2-1-3-5
Frequency	1	1	1	1	1	2	1
Move sequence	2-3-4	2-3-4-5	2-3-5	2-4	2-4-3-4	2-4-5	3-4-5
Frequency	1	9	3	1	1	2	1

A total of 28 move sequences are counted in the 80 abstracts of Chinese EFL learners, among which the move sequence M1-M2-M3-M4-M5 appears the most times, reaching 19 times, accounting for 23.75%, and indicating that Chinese EFL learners' abstracts are roughly in line with Hyland's move model and Chinese EFL learners have a certain awareness of move structure. Besides, there are omitted move, combined move, repeated move, and reversed move in the other 27 move models. The 28 move sequences are counted, in which the case of the omitted move occurs 61 times (76.25%), the case of the combined move occurs 39 times (48.75%), the case of the repeated move occurs 6 times (7.5%), and the case of the reversed move occurs 8 times (10%). From the above data, it can be concluded that the phenomenon of omitted move and

combined move occur more frequently in the abstract. Some Chinese EFL learners write repeated moves, which deserve attention.

The existence of unconventional moves can achieve special communicative purposes. For example, in unconventional moves, moves with repetition or in reverse order can clarify the content and attract readers' attention to important concepts. When the content can be appropriately combined and reduced, the omission and combination of moves can make the text more concise, so that the content is properly handled in the places where it should be detailed and abbreviated, and more focused. The moves of the author's abstract of the test's paper also reflect the omission of moves. Its move sequence is (M1-S4), (M3-S6, S7, S8) (M4-S9), and (M5-S13), omitting M2 (stating the general and/or specific purpose of the study, including hypotheses).

The discussion of unconventional move structure could point to a new analytical perspective for teaching second language writing. It is also beneficial for the study of English abstract writing of research articles. Having reviewed the research on unconventional move structure, this study concludes that it is both practicable and necessary to break free from the research article abstract's general theoretical writing form and employ some appropriate unconventional move patterns in order to better achieve the research article abstract's communicative purpose.

5.2 Analysis of Textual Competence

Based on a text analysis of 80 Chinese EFL learners' abstracts in three aspects (i.e., sentence pattern, tense, and voice), text errors are counted and their types are summarized the overall situation of textual competence. In addition, among the 80 abstract texts, two abstracts showed the list of keywords, which could also be regarded as a manifestation of textual awareness.

5.2.1 Sentence pattern

There are four sentence patterns in English, namely, simple, complex, compound, and compound-complex sentences. In this part, the frequency and percentage of each sentence pattern in each move would be discussed (as shown in Table 10).

Table 10: The Frequency and percentage of the four sentence patterns in each move in the abstracts of Chinese EFL learners

Occurrence of move sequence							
Move sequence	1-2	1-2-3	1-2-3-4	1-2-3-4-1	1-2-3-4-5	1-2-3-5	1-2-3-5-3
Frequency	3	2	3	1	19	6	1
Move sequence	1-2-4-5	1-2-5	1-3	1-3-2-4	1-3-4	1-3-4-5	1-3-4-5-2
Frequency	2	5	1	1	2	7	1
Move sequence	1-3-5	1-4-5	1-5	2-1-2	2-1-3-4	2-1-3-4-5	2-1-3-5
Frequency	1	1	1	1	1	2	1
Move sequence	2-3-4	2-3-4-5	2-3-5	2-4	2-4-3-4	2-4-5	3-4-5
Frequency	1	9	3	1	1	2	1

(F=frequency; P=percentage; T=total)



From the statistics in the table, we can see that: in terms of sentence patterns, simple and complex sentences are the two most used sentence types in Chinese EFL learners' abstracts, while compound and compound-complex sentences are less used. In terms of the number of valid sentences counted, M1, M2, and M5 account for more sentence types, while M3 and M4 are less frequent, which can indicate to a certain extent that Chinese EFL learners describe M1, M2, and M5 more and M3 and M4 less.

5.2.2 Sentence pattern

Four tenses are tallied in the abstracts of Chinese EFL learners after examining the tenses of the abstracts: the simple present, the simple past, the simple future, and the present perfect tenses. This investigation examines and evaluates the distribution of these four tenses in the five moves, as well as their purposes. The results are shown in Table 11.

Table 11: Tense of each move in the abstracts of Chinese EFL learners

Tense type	M1		M2		M3		M4		M5		Total
	F	P	F	P	F	P	F	P	F	P	
Simple present	67	30.73%	39	17.89%	27	12.39%	38	17.43%	47	21.56	218
Simple past	1	2.04%	5	10.20%	24	48.98%	14	28.57%	5	10.20%	49
Simple future	0	0	8	38.10%	2	9.52%	0	0	11	52.38%	21
Present perfect	9	69.23%	0	0	3	23.08%	0	0	1	7.69%	13
Total	77	25.58%	52	17.28%	56	18.60%	52	17.28%	64	21.26%	301

(F=frequency; P=percentage)

According to the data in the table, in Chinese EFL learners' abstracts, the simple present tense is most commonly employed in all moves, whereas the simple past tense, simple future tense, and present perfect tense are rarely utilized. In other words, Chinese EFL learners generally prefer to write abstracts in the simple present tense, probably because this tense is usually used to state facts. The simple present tense dominates the M1 writing in all content, as can be seen in the introduction where it is used many times to state the definition of terms, introduce background knowledge, point out research gaps, and introduce the focus and features of the study.

For the five moves, the simple present tense is most often used to introduce M1 and M2, because it shows the objectivity of the content, helps introduce the topic, and allows the reader to accept the information. In M3, the simple past tense is the main tense as well as the simple present tense. The simple past tense here serves to illustrate the analysis and experimental procedures to briefly recapitulate the research process. In M4, the simple present tense is still the most prominently used tense, and most Chinese EFL learners choose it to demonstrate what they find. This result is consistent with Salager-Meyer's claim that the simple present tense, especially in the commentary type of discourse where conclusions, recommendations, and data are synthesized, serves to reinforce and emphasize the generalizability of specific findings. The dominant tense in M5 is still the simple present, but there is another tense worth noting: the simple future tense, which 14% of Chinese EFL learners tend to use to express the impact of this study and their expectations for subsequent studies.

5.2.3 Voice

Voice is a controversial linguistic feature in abstract writing. Some researchers (e.g., Graetz, 1985; Swales, 1990) believe that passive voice should be used in abstract writing, while others (e.g., Liu & Zhang, 2015) hold the opposite view. In this part, the voice of each move in the Chinese EFL Learners' abstract is as follows.

Table 12: Voice of each move in the abstracts of Chinese EFL learners

Voice type	M1		M2		M3		M4		M5		Total
	F	P	F	P	F	P	F	P	F	P	
Active voice	51	22.08%	49	21.21%	36	15.58%	42	18.18%	53	22.94%	231
Passive voice	25	35.71%	5	7.14%	20	28.57%	9	12.86%	11	15.71%	70

(F=frequency; P=percentage)

As shown in Table 12, the use of active voice far exceeds the use of passive voice in all moves in Chinese EFL learners' abstracts. Using the active voice would be more concise and powerful. This contributes to conveying the research information with clarity and engaging the attention of whom read the lines.

5.2.4 Results of text errors

In the 80 Chinese EFL learners' abstracts, a total of fifty-seven textual errors are counted.

First, there are grammatical errors. In terms of words, there are fourteen misspellings of words, eighteen verbs with the wrong form, three improper verb collocations, and one incorrect use of the article. Seven syntax errors in sentences are found.

Secondly, there are errors in the content. Six cases of a serious imbalance in detail and eight cases of obvious repetition of narrative are found.

The above statistics may conclude that Chinese EFL learners have certain deficiencies in their textual ability and need to improve their ability of standardized writing.

5.3 Analysis of Social Competence

Analysis of social competence is conducted by using Hyland's classification criteria to count and analyze the disciplinary identity discursive markers of Chinese EFL learners' abstracts.

5.3.1 Positioning and Proximity of disciplinary identity discursive markers

The persuasive power of academic texts is not only reflected in the selection of appropriate utterance structures according to the discursive paradigm of the disciplinary community or the strategic use of the resources of the genre, but more importantly, it is inseparable from the author's display of an appropriate disciplinary identity between the lines.

According to Hyland's discursive markers of positioning and proximity in disciplinary identities, 321 cases of identity expressions are found in Chinese EFL learners' abstracts, with 4.0125 cases per abstract. Among them, 149 cases of positioning expressions and 172 cases of proximity expressions are found.

Table 13: The Frequency and percentage of discipline identity discursive markers in the abstracts of Chinese EFL learners

Markers	Text	Text (per 100 words)	Percentage	Rank
Positioning	149	1.77	46.42%	
① Hedges	23	0.27	7.17%	4
② Boosters	7	0.08	2.18%	7
③ Attitude markers	93	1.10	28.97%	2
④ Self mention	26	0.31	8.10%	3

	172	2.04	53.58%	
Proximity				
① Transitions	129	1.53	40.19%	1
② Frame markers	21	0.25	6.54%	5
③ Endophorics	17	0.20	5.30%	6
④ Evidentials	4	0.05	1.25%	8
⑤ Code glosses	1	0.01	0.31%	9

The above data show that compared to other studies (Jiang, 2021), Chinese EFL learners' overall use of disciplinary identity discursive markers in abstract writing is less frequent and their social contextual competence needs to be improved.

Specifically, Chinese EFL learners use proximity slightly more than positioning, which may indicate that they perform better in constructing disciplinary community membership than in manifesting their discursive voice and position. In establishing proximity relationships, Chinese EFL learners focused mainly on transitions and on expressing semantic relationships between clauses through the use of logical connectives. In expressing positioning, Chinese EFL learners focused on attitude markers, expressing the author's attitude clearly and affirming the importance and significance of the study.

5.4 Discussion

5.4.1 Discussion of Generic Competence

In terms of moves, Chinese EFL learners' abstracts include all the typical moves, with no specific move being mandatory. M1, M2, M3, M4, and M5 are all standard moves, and there is no compulsory move to be included. Most Chinese EFL learners tend to follow a four-move model for their abstracts. The most frequently used move in Chinese EFL learners' abstracts is M2, while M4 is the least used. These findings are consistent with Hyland's (2000) five-move model, which suggests that research abstracts generally adhere to a structured format, though variations do exist. However, Chinese EFL learners' awareness of presenting research results needs further development, as the overall structure of moves often lacks logical coherence, a finding echoed by Jiang (2012), who noted that Chinese EFL learners often struggle with the proper structuring of academic texts.

Regarding steps, all the steps in the classification are included in the Chinese EFL learners' abstracts, except for S12, which involves stating the limitations of the research. Among these steps, there are no mandatory ones. Regular steps, such as S1, S5, and S9, show that Chinese EFL learners predominantly focus on arguing the importance of the topic, describing the study's purpose, and outlining the main findings. This supports the findings of Bhatia (1993), who emphasized the importance of defining research objectives and contributions in academic writing. Optional steps, including S2, S3, S4, S6, S7, S8, S10, S11, and S13, suggest that Chinese EFL learners have the flexibility to include various other aspects in their abstracts, such as defining the topic, identifying the research gap, introducing participants, research tools, and research procedures, and analyzing the significance of the research in conclusions and recommendations.

Regarding move sequences, the abstracts of Chinese EFL learners cover all the moves in the classification, but they exhibit considerable variation in move sequences. A total of 28 different move models were identified, indicating substantial variability. Although the 80 Chinese EFL learners' abstracts generally adhere to the five-move model and show some understanding of move structure, we also observed unconventional phenomena such as omitted moves, combined moves, repeated moves, and reversed moves. This finding is in line with studies by Swales (1990) and Hyland (2000), who identified similar deviations in move sequences, suggesting that such variations are common in academic writing across cultures.

5.4.2 Discussion of Textual Competence

Regarding sentence type, simple and compound sentences are the two most common sentence types in Chinese EFL learners' abstracts, while complex sentences are used less frequently. This finding is consistent with the observations of Wang & Slater (2016), who found that Chinese EFL learners tend to rely on simpler sentence structures.

As for tense, the simple present tense is the most frequently used across all moves in Chinese EFL learners' abstracts. In comparison, the simple past tense, simple future tense, and present perfect tense appear less often. This aligns with the findings of Hyland (2002), who noted that the present tense is commonly used in abstracts to present general facts and research findings.

In terms of voice, the active voice is much more commonly used than the passive voice in the abstracts of Chinese EFL learners. This is consistent with recent studies by Wang (2018), who highlighted the preference for active voice in Chinese academic writing.

Regarding textual errors, a total of 60 textual errors were identified in the abstracts of the 80 Chinese EFL learners. Compared to studies by scholars such as Jiang and Cai (2010), this error rate is higher, indicating some weaknesses in textual competence. These findings suggest that while Chinese EFL learners are generally able to construct abstracts, they still face challenges in mastering the syntactic and grammatical aspects of academic writing.

In summary, Chinese EFL learners predominantly use simple sentences, compound sentences, simple present tense, and active voice, but their abstracts contain a significant number of grammatical errors and inappropriate content. This supports the conclusions of Xie (2017), who found similar issues with grammatical errors in Chinese EFL writing.

5.4.3 Discussion of Social Competence

Analysis of the use of discursive markers related to disciplinary identity and proximity in Chinese EFL learners' abstracts may indicate the following patterns.

Firstly, Chinese EFL learners generally use fewer discursive markers of disciplinary identity in their abstract writing, indicating a need to improve their social contextual competence. This finding supports the work of Paltridge (2001), who argued that establishing a clear disciplinary identity is crucial for academic writing.

Secondly, proximity markers are used slightly more often than positioning markers, suggesting that Chinese EFL learners are more focused on establishing membership within the disciplinary community than on expressing their personal stance. This observation is consistent with studies by Hyland (2000), who found that academic writers often prioritize affiliation with the disciplinary community over personal positioning in their writing.

Thirdly, in establishing proximity, Chinese EFL learners mainly emphasize transitions and use more logical connectives. In terms of positioning, they focus primarily on attitude markers, which help express their views and underline the importance and significance of the study.

In conclusion, academic authors must establish proximity with their disciplinary community while expressing their own positioning. This balance is crucial for effectively constructing their disciplinary identities. Thus, Chinese EFL learners need to increase the use of disciplinary identity markers and strengthen their expression of positioning to better balance proximity and positioning, thereby enhancing their social competence in academic writing. These recommendations are supported by the findings of Hyland and Tse (2004), who emphasized the importance of academic identity and positioning in scholarly communication.



6 Conclusion

Based on Jiang's (2021) theory of academic discursive competence, this study analyzes the abstract writing of Chinese EFL learners in research articles across three levels: generic competence, textual competence, and social competence. At the level of generic competence, it is evident that Chinese EFL learners are somewhat aware of move structures, but their organization of moves lacks coherence. While most learners opt for the four-move structure, there are notable inconsistencies in the sequence of these moves, including omissions, combinations, repetitions, and reversals of moves. These unconventional patterns suggest a need for further improvement in structural awareness.

Regarding textual competence, Chinese EFL learners demonstrate some understanding of textual features; however, they struggle with maintaining grammatical accuracy, particularly with vocabulary and sentence structure. Their abstracts predominantly employ simple sentences, complex sentences, the simple present tense, and active voice. However, a considerable number of grammatical errors and inappropriate content still appear in their writing.

At the level of social competence, three major observations were made. First, Chinese EFL learners generally use fewer disciplinary identity markers in their abstracts, indicating that their social competence requires further development. Second, they focus more on establishing membership within the disciplinary community rather than expressing their individual discursive voice and position. Third, they primarily use logical connectives in transitions and focus on expressing attitudes through attitude markers. This suggests that learners are more concerned with connecting ideas than articulating their academic stance.

This study has several limitations that could be addressed in future research. First, the sample size of 80 abstracts is relatively small and may not fully represent the broader population of Chinese EFL learners. These abstracts were also sourced exclusively from senior undergraduate English majors at double-class universities in central China, which may limit the generalizability of the findings. Future studies should include a larger and more diverse sample of texts to improve representativeness. Second, while this study focuses on analyzing the characteristics of Chinese EFL learners' abstract writing, it does not systematically compare these abstracts with those written by the original authors of the research articles. Future research could address this gap by conducting a comparative analysis of Chinese and foreign abstracts, focusing on differences in structure, linguistic features, and rhetorical strategies, and exploring the reasons behind these discrepancies. The findings suggest evidence-based pedagogical strategies that could enhance Chinese EFL learners' academic discursive competence and contribute to more effective academic writing instruction. This study highlights the need for Chinese EFL learners to improve their engagement in international academic contexts by developing competence in standardized, discipline-specific texts, understanding the genre's clear move-step structure, and mastering the social context of academic communities. Improving academic discursive competence appears to be essential for cultivating internationally competitive talents.

This study contributes to raising awareness of the challenges faced by Chinese EFL learners in academic writing and points to the importance of developing a stronger academic identity, which may enable learners to engage more effectively in global academic communication. It contributes to the cultivation of a robust academic selfhood, potentially empowering learners to interact confidently with international readers and academic communities.

From a pedagogical perspective, the findings suggest that universities and instructors should consider tailoring academic English instruction to meet the diverse needs of different disciplines and EFL learners. By offering context-specific training and feedback, academic English teaching could be better adapted to address the challenges faced by Chinese EFL learners. Additionally, higher education institutions in China might need to recognize the importance of academic English education in preparing students for global academic exchange and could allocate sufficient resources to support this area of language learning.

REFERENCES

- Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. Longman Press Publishing.
- Bhatia, V. K. (2004). *Worlds of written discourse: A genre-based view*. Continuum.
- Cai, J. G. (2019). Reading and writing research articles of international journals among Chinese undergraduates: Move identification and analysis. *Foreign Language Learning Theory and Practice*, (3): 76-82.
- Cai, J. G. (2022). Writing SCL journal Research Articles for Publication. *Academic Degrees & Graduate Education*, (2): 43-50.
- Canagarajah, S. (2002). *A geopolitics of academic writing*. Pittsburgh: University of Pittsburgh Press.
- Canagarajah, A. S. (2013). *Critical academic writing and multilingual students*. University of Michigan Press.
- Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative analysis. *International Journal of Qualitative Studies in Health and Well-being*, 1(3). <https://doi.org/10.3402/qhw.v1i3.4932>
- Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.
- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Sage.
- Geng, et al. (2024). Analyzing engagement strategies in argument chain: A comparison between high- and low-scoring EFL undergraduate argumentative essays. *Journal of English for Academic Purposes*, 71, Article 101428.
- Graetz, N. (1985). Teaching EFL students to extract structural information from abstracts[J]. *English for Specific Purposes*, (24):123-267.
- Huang, Y.Y., Jia, H. & Li, D.C. (2025). Transvocal stance in academic translation: A rhetorical analysis of grammatical stance in translated applied linguistics English research article abstracts. *Journal of English for Academic Purposes*, 74, Article 101472.
- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6, 183-206.
- Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. Longman Press Publishing.
- Hyland, K. (2004). *Disciplinary Discourses: Social Interactions in Academic Writing*. Ann Arbor, MI: Michigan University Press.
- Hyland, K., & Tse, P. (2004). Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics*, 25, 156-177. <http://dx.doi.org/10.1093/applin/25.2.156>
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192. <https://doi.org/10.1177/1461445605050365>
- Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse*. Cambridge University Press.
- Hymes, D. (1972). On Communicative Competence. In J. B. Pride, & A. Holmes (Eds.), *Sociolinguistics: Selected Readings*. Harmondsworth: Penguin.
- Jiang, F. (2021). The Development of Chinese Students' Academic Discursive Competence and the Effect of Teaching. *Foreign Languages and Their Teaching*, (6): 34-44+148.
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles[J]. *English for specific purposes*, 24(3): 269-292.
- Kurniawan, E., Lubis, A. H., Suherdi, D., & Danuwijaya, A. A. (2019). Rhetorical organization of applied linguistics abstracts: does Scopus journal quartile matter?. *GEMA Online Journal of Language Studies*, 19(4), 184.



- Liu, Y. H. & Zhang, Y. (2016). A contrastive study of the abstracts in international journal articles by Chinese and English-speaking scholars. *Foreign Language World*, (5): 20-27.
- Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge University Press.
- Paltridge, B. (1996). Genre and the language learning classroom. *Elt Journal*, 50(3), 598-635.
- Rowley, J. (1988). *Abstracting and indexing*. Clive Bingley.
- Salager-Meyer, F. (1990). Discoursal flaws in medical English abstracts: A genre analysis per research-and text-type. *Text-Interdisciplinary journal for the study of discourse*, 10(4), 365-384.
- Santos, M. B. D. (1996). The textual organization of research paper abstracts in applied linguistics. *Text & Talk*, 16(4): 481-500.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Sui, Y. Z. (2017). Liberal English Education. *Foreign Language Teaching and Research*, 2017, 49 (06): 859-870.
- Wang, J. (2013). Understanding the Chinese Learners from a Perspective of Confucianism. In: Cortazzi, M., Jin, L. (eds) *Researching Cultures of Learning*. Palgrave Macmillan, London. https://doi.org/10.1057/9781137296344_4
- Wang, S., & Slater, T. (2016). Syntactic Complexity of EFL Chinese Students' Writing. *English Language and Literature Studies*, 6(1), 81. doi:10.5539/ells.v6n1p81
- Xie, Q. (2017). Diagnosing university students' academic writing in English: Is cognitive diagnostic modeling the way forward? *Educational Psychology*, 37(1), 26-47.
- Zhao, N. (2023) A Corpus-based comparison study of first-person pronoun we in English-language abstracts. *Journal of English for Academic Purposes* (online).
- Zubir, F. , Suryani, I., & Abdullah, S. (2021). A comparison study on the rhetorical moves of abstracts in engineering master and PhD best thesis. 2339(1): 020052.

Examination of the Prevalent Biases in Digital Composition of Chinese Characters

Shubin Li

School of Foreign Languages, Tianjin Chengjian University, Tianjin 300384, China

Abstract

The present study meticulously reviews the literature concerning the biases in “electronic scripting of Chinese characters” from 2016 to 2024. It synthesizes the research findings, methodologies, and conclusions from prior inquiries into the biases associated with electronic scripting. The review indicates a primary focus within the research community on the taxonomic categorization, determinants, and pedagogical strategies, with the methodology targeting predominantly the categorization system and the influencing factors. The analysis reveals that the focus of research is centered on the categorization system, determinants, and pedagogical strategies, with a preference for a combination of quantitative and qualitative methodologies. The conclusions of these studies have evolved from straightforward typological analysis to an exploration of underlying and effective teaching approaches. Nevertheless, the sample population in these studies lacks generalizability, the impact of cultural context on students’ biases in “electronic writing mechanisms” has not been thoroughly investigated, and there is a deficiency in longitudinal tracking of learner’s progress over time. To enhance the applicability of the findings to a broader demographic, future research could incorporate questionnaires, interviews, and longitudinal studies.

Keywords: e-writing; Chinese characters; errors; second language acquisition; teaching strategies

1 Introduction

With the development of multimedia, the status of intelligent equipment in teaching is increasing, and electronic learning (electronic Writing, e-writing) came into being, and gradually became a new model and new demand in language acquisition and teaching (Xie, 2008, 2014). Entering the post-epidemic era, simple handwritten Chinese characters can no longer meet the requirements of teachers and students for learning efficiency. Many scholars predict that “paperless” learning will become a general trend (Yang and Zhan, 2024:52). In this context, the hypothesis that input methods promote Chinese character acquisition has been repeatedly mentioned, but early studies only focused on the auxiliary functions of the electronic writing mode and failed to fully explore its application value in acquisition. This limitation triggered the academic community to discuss the new paradigm of “electronic writing”, and began to compare the differences in the implementation of handwritten Chinese characters and electronic writing Chinese characters. The focus has gradually shifted to “how to promote acquisition”. In 2022, several cloud conferences held on the theme of “Chinese Character Teaching in the Electronic Writing Era” are not only a milestone in the internationalization of the electronic writing model, but also the first large-scale attempt to combine electronic writing theory with practice. The strategy of “electronic writing as the main method and handwriting as the auxiliary method” (Zhai, 2024; Chu, 2024) emphasizes the core role of electronic writing in Chinese character acquisition and teaching. While it has been widely recognized, it also marks the beginning of the systematic and in-depth development of research in this field, providing an empirical basis for future theoretical frameworks.

Different from the observation channel of handwriting errors, errors in electronically written Chinese characters focus more on the errors caused by emerging electronic products (such as mobile phones and computers) in the process of Chinese character input and recognition. Xie (2005) of California State University, Long Beach (CSULB) has analyzed the errors and causes of Chinese phonetic input since the beginning of the century. In the past decade, the field of Chinese teaching has gradually used electronic writing and typing as the main teaching method. Related research has begun to focus on the errors caused by electronic writing of Chinese characters, digging deep into its mechanism and building a corresponding theoretical system to meet application needs. After a comprehensive analysis of existing literature, it is found that previous studies have explored errors in electronic writing of Chinese characters from multiple dimensions and constructed a preliminary framework that covers the types of errors, influencing factors and teaching strategies. These articles are mainly from universities such as Beijing Language and Culture University, Guangdong University of Foreign Studies, Vassar University, and Michigan State University. The keywords revolve around “spelling”, “electronic writing”, “errors”, “teaching” and “students”, revealing the new characteristics of Chinese character acquisition in the electronic writing environment.

In addition, the research of Chen (2016) and Jiang (2017) showed that there is a significant correlation between electronic writing input and Chinese character acquisition. On the one hand, electronic writing can promote Chinese character processing and acquisition; on the other hand, electronic writing of Chinese characters can not only accelerate the process of Chinese character acquisition, but also provide an effective window for observing the error mechanism.

2 Literature Review

“Electronic writing” refers to the process of typing Chinese characters on a computer (or mobile phone) or inputting Chinese characters by voice, and displaying Chinese characters on the screen without “writing Chinese characters” (Lu, 2024: 3). From the name change from “spelling” to “electric typing” and then to “electric writing”, we can observe the evolution of Chinese character input technology and the innovation of teaching practice. At first, the term “spelling” mainly referred to the fact that Chinese character input based on the pinyin input method must be completed through two steps: phonetic input and glyph recognition. With the development of science and technology, people began to pay attention to the fact that Chinese

character input is not only a matter of means, but also the interaction between equipment, platform and people. Therefore, the term “electric writing” came into being. It not only includes the operational characteristics of spelling, but also includes the importance of equipment in input. It also emphasizes the handwriting input and voice input brought about by technological updates (Shi and Ji, 2024), and at the same time points out the development path from word acquisition to discourse acquisition.

Regarding the definition of electronic writing errors, different scholars have proposed different definitions from their respective research perspectives. Chen (2016) mentioned that electronic writing errors mainly refer to the errors made by international students when using smart electronic products to input Chinese characters, reflecting the common errors in Chinese character recognition and input in the electronic writing environment. Although Jiang (2017) did not directly define electronic writing errors, he proposed the combination of handwriting and electronic writing input, which also hinted at the relationship between handwriting errors and electronic writing errors. Che (2023) summarized electronic writing errors as Chinese character errors caused by students’ cognitive differences or operational errors in the process of spelling Chinese characters. Combining the above viewpoints, we can give a more comprehensive definition of electronic writing errors: electronic writing errors are errors made by students when inputting Chinese characters in an electronic writing environment. The errors involve multiple levels of cognition, means, and equipment, reflecting students’ deficiencies in Chinese character cognition and application.

3 Research Design and Methods

3.1 Type research

The existing electronic Chinese character error system can be roughly divided into four categories: pronunciation errors, character shape errors, input method association errors and other errors.

3.1.1 Pronunciation errors

In the literature on errors in electronic writing, pronunciation as a key point covers errors in finals, initials and “er”. Its complexity lies in the variety of types and high difficulty in judgment (Chen, 2023; Zhai, 2024), which is particularly common in the primary stage. Among them, errors in finals are significant and often associated with pitch differences, specifically manifested as confusion between front and back nasal sounds. Thanks to the uniqueness of electronic writing input, the error mechanism of the erhua final “er” is highlighted: studies have found that “er” is often typed with the previous syllable as r, resulting in confusion between “儿” and “人”, or it cannot be typed without being restored to “er” and is simply “abandoned” by students (Chen, 2016; Chen, 2013). These phenomena also provide a new perspective for examining the special pinyin mechanism of error correction. Compared with errors in finals, errors in initials often involve errors in sound length and pronunciation position, such as confusion between front and back sounds at the tip of the tongue (Chen, 2016). The above fully reflects the physiological characteristics contained in acquisition errors. In addition, the cognitive errors pointed out by physical features are not limited to oral learning. For example, the substitution of words or syllables with similar pronunciations also occurs in electronic writing environments. It is worth noting that these errors do not exist in isolation. They are closely related to the age group, native language background, and level of the learners.

3.1.2 Font errors

Turning to the shape error, it is the most frequent type of all errors, including errors of homophones with similar shapes, homophones with different shapes, and heterophones with different shapes (Zhai, 2024). In particular, homophone/word errors account for a considerable proportion in later learning, and the distribution characteristics also echo the error pattern of handwritten Chinese characters. Pinyin dependence, as a common phenomenon among beginners, can expose the shortcomings of shape recognition. Once students are required to “type by looking at the characters” without the assistance of pinyin, the shape substitution



errors will increase significantly, but in comparison, the shape errors are still effectively controlled in the “electronic writing” environment (Chen, 2016). Turning to the structural perspective, shape errors are further broken down into component errors and planar diagram errors (Chen, 2023), that is, component combination errors of compound characters. Compound characters are often “deleted” by students into single characters, and the phenomenon of “mutual substitution and mixing” between different radicals also often occurs. Whether the error characters are familiar characters or new characters, students will be confused. In his research, Zhai (2024) creatively paid attention to the confusion between simplified and traditional Chinese characters, which mainly exists among overseas Chinese students with a Chinese character background. This finding provides a sample for exploring the complex mechanism of electronic writing errors.

3.1.3 Input method errors

This type of error is unique to electronic Chinese writing. Common errors include misleading associations, incorrect input of connected sentences, and missing empty characters. All of the above phenomena reflect that students have an avoidance mentality in the learning process (Chen, 2016). Specifically, elementary-level students are unable to be compatible with the intelligent association function of the input method. When inputting, they may be influenced by the system’s recommendations and choose inappropriate vocabulary or phrases. In addition, incorrect input of connected sentences also occurs from time to time. Some students will input a whole sentence at a time without careful proofreading, resulting in confusion of meaning groups or grammatical errors. In fact, this is also a sign that students are overly dependent on the input method. Finally, most of the missing empty characters are because students fail to master the target characters and choose to use spaces as placeholders. A small part is due to accidental keyboard touches, especially when the virtual keyboard is not strong enough to operate, it is more likely to produce such errors (Spilling, Rønneberg, Rogne, Roeser, Torrance, 2022). However, it is speculated that it may also be because European and American students are influenced by the English input habit and subconsciously put spaces between characters or words, resulting in missing empty characters.

3.1.4 Other biases

These errors are relatively trivial, with an irregular frequency of occurrence, and they present different forms at different stages. For example, semantic errors and grammatical errors are common among young students in the initial stage of learning. Students have not yet mastered the rules of the target language or overgeneralize them, and use participles inappropriately, omit Chinese characters, or use redundant phrases (Zhai, 2024). According to Chen’s (2023) study of college students in the initial stage of learning, phoneme addition and subtraction errors are even more variants of the phenomenon of overgeneralization, which indirectly reflects the difficulties students have in acquiring phonetics. Overgeneralization is not limited to the lexical level but also reflected in grammatical structures, such as the loss of medial, which together constitute a diverse network of errors in electronically written Chinese characters.

3.2 Research on influencing factors

On the one hand, mother tongue transfer is more common among international students in the initial stage of learning. Because of their first contact with Chinese, students will directly embed their mother tongue thinking mode and even the writing form of their mother tongue into Chinese, resulting in the phenomenon of replacing radicals. At this time, they do not follow the set rules, but spell according to their own habits. Once they encounter unfamiliar Chinese characters in practice, they will replace them with Chinese characters with the same or similar pronunciations. This “replacement” phenomenon is actually a transfer within the Chinese language. On the other hand, there are also positive and negative transfer effects in the cognitive mechanism that affect the acquisition of Chinese characters, and are closely related to the complex process of memory processing. Among them, metalinguistic awareness has a positive impact on the distinction between phonetics, components and compound morphemes (Chen, 2023). Yang (2024) and others have

confirmed that under certain conditions, electronic writing is conducive to reducing the fear of second language students in learning Chinese characters and promoting their Chinese character cognition.

Finally, improper teaching strategies can also lead to errors in electronic writing. Due to the particularity of electronic writing, the errors caused by this influencing factor run through almost every teaching link, and the frequency of occurrence is disordered and the form is difficult to predict. In the initial stage, some teachers overemphasized the accuracy of spelling word by word. This teaching orientation inadvertently prompted students to disassemble compound words and even split and input complex Chinese characters one by one, which gave rise to a large number of literacy and spelling errors. Such errors actually directly reflect the lack of teacher experience and prove that refined strategies are very necessary. In addition, defects in textbook design are also one of the causes of errors (Chen, 2016). At present, there is a lack of clear operational guidelines for easily confused or special pinyin such as “er” and “u”, which undoubtedly increases the difficulty of learning. At the same time, there is a lack of teaching resources based on large-scale corpora, and it is difficult for teachers to foresee the difficulties that students may encounter, and it is difficult to provide guidance and adjustments in a timely manner; if students’ language foundation is not solid, teachers blindly increase the amount of electronic writing tasks, which will only increase the students’ cognitive burden (Zhai, 2024). This “quantity accumulation” rather than “qualitative improvement” directly leads to an increase in the error rate.

3.3 Teaching strategy research

The teaching strategies for errors can be roughly summarized into three dimensions: content, methods, and application. In terms of content optimization, first, we should focus on the recognition of pronunciation and design various exercises to consolidate the connection between pronunciation, shape, and meaning of students, so that students can be familiar with the processing mechanism of Chinese characters that is different from that of phonetic characters. Second, we should focus on the key role of components and radicals in the recognition of Chinese characters (Zhai, 2024). At the same time, Chen (2016) also proposed to use the “learning accompanies and using accompanies” phenomenon generated by the context and Chinese characters in the learning process to enhance the interactive effect with students to promote learning; simultaneously create a “word” context, integrate Chinese characters into the local context of the language for learning, and read them multiple times in series with the context to reduce the rate of character recognition errors. In general, there should be less error correction in the initial stage, and rich context information should be reconstructed to “accompany” the learning of Chinese characters to reverse potential teaching errors.

In terms of method innovation, some scholars have proposed to introduce the principle of separating the recognition and writing of handwritten Chinese characters to separate the recognition and writing of electronically written Chinese characters (Jiang, 2017), indicating that breaking down complex tasks to allow students to focus on a single skill can reduce the difficulty of learning. Secondly, the study takes into account the importance of artificial shape recognition awareness and ability for Chinese character learning. It not only follows the path of phonetic training being regarded as enhancing shape recognition ability, but also expands the guiding role of phonetic indicators, which is convenient for urging the integration of example sentence corpora of electronically written Chinese character errors for teaching intervention (Zhai, 2024).

In terms of application, we should focus on the preparation and introduction of electronically written Chinese characters using computers, mobile phones and other electronic devices, and then introduce Chinese character functional components in a timely and appropriate manner in the electronic writing-based mode, keeping handwriting as an auxiliary role in electronic writing. Teachers should also release the teaching function of pinyin and make good teaching designs for pinyin typing.



3.4 Experimental method study

The research on errors in electronic writing is relatively traditional in experimental methods and is in its infancy. Most of the research subjects are Chinese beginners in European and American countries, and the investigation period is relatively short, ranging from one to six months. Most of the selected corpora have been pre-designed teaching links and teacher intervention, and then obtained through homework and daily evaluation. In the data processing stage, the annotation results and the above-mentioned data processing methods were used to ensure the accuracy of the data and the effectiveness of the analysis (Zhai, 2024). In the process, the collected error examples were classified, annotated and statistically analyzed, and the common error types and causes of students in the process of learning Chinese were analyzed.

4 Comments on Previous Studies

4.1 Review of type studies

In the study of types, previous studies have focused on introducing different dimensions and comparing with traditional Chinese character errors, constructing a more intuitive error type system, highlighting the general mechanism of Chinese character errors and the unique characteristics of electronically written Chinese characters. And taking into account the special errors caused by electronic writing equipment. However, the disadvantage is that the inductive boundary of errors is vague. Many special errors have not been discussed separately, but are only unilaterally classified from the perspective of shape and sound. For example, previous studies have not discussed whether the voice input link of electronically written Chinese characters will produce tone errors. Tone is a major difficulty for Chinese students to learn Chinese characters (Fang and Ren, 2024), and students in the electronic writing environment do have character shape errors caused by tone. In addition, the erhua sound, as one of the difficulties for Chinese students to learn Chinese characters, has not received enough attention. Its error forms are complex, which reflects that students do not have enough knowledge of morphemes and that they are not very clear about the input principle of electronic writing equipment.

4.2 Review of research on influencing factors

Based on the current research on the factors affecting electronic writing errors, it can be found that the interdisciplinary perspective is more prominent. Researchers have analyzed the principles of electronic writing errors from multiple dimensions such as linguistics, psychology, and education, showing a dynamic and progressive view. The shortcomings are that the causes of some errors are only inferred based on morpheme knowledge, and students are rarely asked why they miss or make mistakes when typing; the research on the path of mother tongue transfer is not in-depth. Taking beginners in Europe and the United States as an example, it does not consider how English letter combinations affect the occurrence of errors such as confusion and omissions in electronic writing of Chinese characters. In addition, it does not consider whether there is a Chinese character background, different teaching scenarios, and various types of Chinese texts will present what kind of error types and their characteristics and trends.

4.3 Review of research on teaching strategies

Previous studies on teaching strategies for electronic writing errors have consciously introduced an interdisciplinary perspective and fully considered the relationship between the brain and electronic writing errors (Cui, 2024; Yang and Zhan, 2024). Strengthening the structure of electronic writing Chinese characters, improving the accuracy of recognition and memory, also enhances students' ability to apply Chinese characters in practice. At the same time, the study also realized the connection between platform interactivity and students' sense of participation and motivation (Zhai, 2024), actively introduced the teaching principles of handwritten Chinese characters to assist electronic writing learning, and set up localized teaching strate-

gies that can be adjusted at any time according to students' specific needs. However, there are also some incomplete considerations. For example, building an effective context and innovating teaching methods may require additional teaching resources, more professional training and teaching design capabilities, which is a challenge for schools and teachers and may also increase the cost of education. In addition, relying on teaching strategies to suppress errors is still dominated by teachers, and it does not focus on what students can do in the process of discovering and solving errors, and guiding the compilation of teaching materials is not on the agenda. More importantly, as electronic devices are a key part of electronic writing, their performance differences may affect learning outcomes. How to solve the accessibility and consistency issues in different regions and under different economic conditions also needs further discussion.

4.4 Review of experimental methods research

Different from the error research of traditional Chinese character acquisition, the error research of electronic Chinese characters is mainly quantitative, and is supported by empirical evidence and data. The corpus is collected through daily practice, which is authentic and reliable. Moreover, the collection is not limited by time and space, and the forms are diverse, which makes it possible to highlight the new errors that have not been discovered in the traditional handwritten Chinese character error research, and facilitates the exploration of the relationship between modern technology and Chinese character teaching. What needs to be paid attention to is that the means selected in qualitative research are relatively simple, and the root causes of students' errors are rarely explored through observation, questionnaires, conversations, etc., and there is no long-term follow-up survey of students to observe the changes in error types with learning progress. It only describes its appearance through case and comparative analysis methods, and the collected corpus does not indicate which type of examination questions are used to obtain it. The researchers have no way to judge whether the error types and influencing factors are objective in the analysis, and the correspondence between each error and the Chinese proficiency stage is poor. In addition, the language background, age group, and level stage of the research subjects are relatively simple, and it is impossible to explore the relationship between these three and the error types. The collected corpus does not specify the type of examination questions used to obtain it. The researchers have no way of judging whether the analysis of error types and influencing factors is objective, and the correspondence between the various errors and the Chinese proficiency levels is poor.

5 Conclusion

After an in-depth analysis of the existing literature, the following conclusions were drawn: First, the types of errors in electronically written Chinese characters are mainly pronunciation errors, character shape errors, input method errors, and other errors. The types show stage characteristics, and the structural system is clear but not very targeted. Second, the influencing factors of electronic writing errors include mother tongue transfer, cognitive processing, and improper teaching strategies. Previous studies have adopted an interdisciplinary perspective but have poor depth and tendency. Teaching strategy research is roughly divided into three dimensions: content, method, and application. It focuses on practicality and memorization efficiency, but the participation of other subjects in the teaching process is not high. Third, the experimental method prefers to collect corpus in advance and then annotate the comparison results. The quantitative support is strong, but the universality and objectivity need to be observed. Future research should be committed to expanding the country diversity of samples, focusing on analyzing new types and new forms of electronically written Chinese character errors, especially those special errors caused by electronic writing interface interactions. At the same time, mixed experiments should be designed to combine quantitative data with qualitative interviews to reveal the deep reasons behind electronic writing errors, design more refined teaching strategies, and enhance the application value of electronically written Chinese characters in reality.



CRedit authorship contribution statement

Ms. Shubin Li finished the study solely.

Declaration of Generative AI and AI-assisted technologies in the writing process

The author declares that no Generative AI or AI-assisted technology was used in the study.

Declaration of competing interest

The author declares that she has no competing interests.

Data availability

The datasets generated and analyzed during the current study are available from the author on reasonable request.

References

- Chen, M. (2023). Analysis of the types, characteristics and influencing factors of Chinese character spelling errors made by American elementary learners. *Chinese Character Culture*, (1): 141-147.
- Chen, Y. (2016). Error analysis and teaching strategies of “spelling and typing” Chinese characters: zero-starting Chinese character teaching based on smart electronic products. *Chinese Learning*, (1): 84-94.
- Chu, C. (2024). Developing Chinese second language teaching through the teaching transformation of “electronic writing as the main method”. <https://b23.tv/kaT1NOE>, 2024-08-01/2024-10-13.
- Chu, C., Liang, N., Lu, X., Qian, Z., Wang, Q., Wang, X., & Zhang, X. (2024). Roundtable discussion on Chinese character teaching in the era of electronic writing. <https://b23.tv/0nOObfL>.
- Cui, X. (2024). The influence of “handwriting” and “electronic writing” on Chinese character learning and related research perspectives. *Journal of Yunnan Normal University (Teaching and Research on Chinese as a Foreign Language)*, 22 (5): 11-19.
- Fang, Y., & Ren, L. (2024). Research on the tone of Chinese learners in Pakistan and teaching suggestions: Taking “bu” and “yi” as examples. *Chinese Character Culture*, (17): 163-168.
- Jiang, X. (2017). Chinese character teaching in the keyboard era. *International Chinese Teaching Research*, (2): 4-10.
- Lu, J. (2024). Typing Chinese characters leads the way, followed by writing Chinese characters: Innovating the Chinese character teaching model in international Chinese education. *Journal of Yunnan Normal University (Teaching and Research on Chinese as a Foreign Language)*, 22 (4): 1-5.
- Shi, F. & Ji, Z. (2024). Electronic writing of Chinese characters is the general trend of the information age. *Journal of Yunnan Normal University (Teaching and Research on Chinese as a Foreign Language)*, 22 (4): 6-11.
- Spilling, E. F., Rønneberg, V., Rogne, W. M., Roeser, J., & Torrance, M. (2022). Handwriting versus keyboarding: Does writing modality affect quality of narratives written by beginning writers?. *Reading and Writing*, 35(1): 129-153. <https://doi.org/10.1007/s11145-021-1069-y>.
- Xie, T. (2014). Chinese teaching keeps pace with the times. *International Chinese*, (1): 69-74.
- Xie, T. (2008). Computer teaching is the only way to make it a daily routine. *Chinese Teaching Modernization Society. Progress and deepening of digital Chinese teaching*. California State University, Long Beach, USA: 7.

- Xie, T. (2005). Error analysis of Chinese Input in Chinese learning. Ping Xu and Theresa Jen Chinese Character Teaching and Computer Technology. Taipei: Department of Education.
- Yang, Q. & Zhan, Z. (2024). “Handwriting” or “typing”? Mechanisms and evidence of the influence of writing style on cognitive learning. *Open Learning Research*, 29 (1): 52-62.
- Zhai, N. (2024). Innovation and practice of Chinese character teaching model in the era of electronic writing: A case study of errors in “electronic writing” by Canadian primary school students. *Journal of Yunnan Normal University (Teaching and Research on Chinese as a Foreign Language)*, 22 (4): 19-28.
- Zhang, N. (2024). L2 New Path for Chinese Character Teaching: Review, Reflection and Prospect of the Effectiveness of Electronic Writing Teaching. <https://b23.tv/dYYTPsk>, 2024-08-01/2024-11-01



A Corpus-Based Comparative Study on the Low-carbon Image Construction of Chinese and American Banks

Xiaoqin Liu¹, Yishan Zhang^{2*}, Zequan Jiang³

¹School of Foreign Languages and Cultures, Guangdong University of Finance, Guangzhou, China

²School of Foreign Languages and Cultures, Guangdong University of Finance, Guangzhou, China

³School of Foreign Languages and Cultures, Guangdong University of Finance, Guangzhou, China

*Corresponding author Email: lisazhang15119839072@gmail.com

Abstract

The global transition to low-carbon development has elevated commercial banks to a pivotal role in advancing carbon finance and national sustainability agendas. China's "dual-carbon" targets (carbon peaking by 2030 and carbon neutrality by 2060) underscore the strategic importance of green finance, prompting banks to leverage carbon finance initiatives for enhancing corporate social responsibility (CSR) images and market competitiveness. This study employs a corpus-based discourse analysis to compare CSR and ESG reports of major Chinese and American commercial banks from 2019 to 2022. It investigates two core dimensions: (1) the low-carbon images projected by banks and (2) the linguistic strategies underpinning these images. By analyzing high-frequency lexical clusters, keyword-in-context (KWIC) patterns, and discursive frameworks, the research reveals distinct cross-cultural narratives: Chinese banks emphasize policy alignment and green financial products (e.g., "green finance," "sustainable development"), while American banks prioritize global climate governance (e.g., "Paris Agreement," "climate risk") and market-led sustainability commitments. The findings offer actionable insights for banks to refine their CSR narratives, aligning with both international climate frameworks and domestic policy imperatives. This comparative analysis bridges the gap between quantitative ESG research and qualitative discourse studies, shedding light on how language shapes corporate environmental identities in global financial sectors.

Keywords: low-carbon image, corpus-based discourse analysis, image construction, Chinese and American banks

1 Introduction

The global shift toward low-carbon development is accelerating, especially as countries pursue climate change mitigation and sustainable development goals. China's 'dual-carbon' targets—carbon peaking by 2030 and carbon neutrality by 2060—highlight the critical role of carbon finance in both national and international financial competition. Commercial banks are central to this transition, as their engagement in carbon finance not only shapes resource allocation but also enhances their market position and brand image. Carbon finance, including green bonds and other financial instruments, is essential for channeling capital into green and low-carbon projects, supporting China's dual-carbon goals and driving sustainable economic transformation (Guo, 2025; Cong, 2023; Hua & Zhang, 2025; Wang et al., 2024; Zhang et al., 2022). Commercial banks are key players in this process, as they provide funding, manage risks, and innovate financial products to support green development (Guo, 2025; Li, 2023; Wang et al., 2024; Xu, 2025). The adoption of green finance by banks promotes technological innovation, supports green industries, and encourages enterprises to pursue low-carbon development paths (Fu et al., 2025; Wang et al., 2024; Xu, 2025). Banks' active participation in carbon finance and green finance initiatives strengthens their corporate social responsibility (CSR) image, enhances competitiveness, and positively influences public perception (Zhang, 2024; Li, 2023).

In the context of corporate social responsibility (CSR), cultivating a low-carbon image has become a crucial indicator of a bank's dedication to environmental responsibility. This image not only highlights proactive sustainability efforts but also enhances public trust and market reputation (Zhang, 2024). However, most existing studies have concentrated on environmental discourse within general corporate settings, leaving a gap in research focused on how financial institutions strategically construct their low-carbon identities through discursive practices (Liu & Zhang, 2024). This research seeks to supplement existing literature by investigating the low-carbon image construction of Chinese and American commercial banks. Through the application of corpus linguistics and discourse analysis, this study examines how these banks articulate their environmental identities in CSR/ESG reports from 2019 to 2022. The research addresses two key questions: (1) What low-carbon image have Chinese and American commercial banks been constructing respectively? (2) How do Chinese and American commercial banks construct their low-carbon image respectively? In other words, the study attempts to understand how language is used by banks to create specific perceptions and associations in the minds of consumers. By analyzing the linguistic techniques utilized by these banks, we can gain insight into the ways in which they craft their public low-carbon image.

The study offers both theoretical and practical contributions. Theoretically, it enhances the interdisciplinary study of corporate environmental discourse by applying corpus-based methods to analyze cross-cultural differences in sustainability communication. Practically, it provides empirical insights that can help banks refine their low-carbon discourse strategies, aligning with global climate governance frameworks such as the Paris Agreement while addressing domestic policy requirements. By comparing discursive patterns between Chinese and American banking sectors, this research also contributes to broader discussions on international low-carbon discourse influence and corporate social responsibility practices.

2 Literature Review

2.1 A Review of Studies on CSR/ESG Reports of Chinese and American Banks

Many studies examine how CSR/ESG disclosures relate to profitability, risk, cost of debt, and stock price crash risk, often using statistical models and financial indicators (Yuan et al., 2021; Hu et al., 2023; Feng et al., 2021; Sun et al., 2022). Some research explores how firm size, ownership, board characteristics, and regulatory environments influence the extent and quality of CSR/ESG reporting (Ervits, 2021; Fu et al., 2023). Some studies use quantitative or mixed-methods content analysis to assess the thematic focus and homogeneity of CSR reports, especially in the context of state-owned versus private banks (Chen et al., 2021;



Ervits, 2021).The effects of ESG ratings and rating disagreements on bank loan availability and market efficiency are also key topics (Hu et al., 2023; Qin & Wang, 2025; Zhang et al., 2024; Sun et al., 2022).The following table in particular presents the research focuses and methodologies of the current studies.

Table 1: The major studies on CSR/ESG reports of banks

Research Focus	Methodology Used	Example Studies
Financial/Market Impact Statistical	Financial/Market Impact Statistical	Tunio et al., 2020; Yuan et al., 2021; Hu et al., 2023; Feng et al., 2021; Sun et al., 2022
Disclosure Determinants	fsQCA/Regression	Ervits, 2021; Fu et al., 2023
Content/Thematic Analysis	Quantitative/Qualitative Coding	Chen et al., 2021; Ervits, 2021
ESG Ratings & Credit Mar- kets	Data Analysis	Hu et al., 2023; Qin & Wang, 2025; Zhang et al., 2024; Sun et al., 2022

As discussed above, most research focuses on quantitative metrics, governance structures, and financial outcomes, rather than the linguistic or discursive construction of low-carbon narratives. Furthermore, existing studies rarely use corpus linguistics methods such as high-frequency words, word clusters, KWIC, or detailed discourse analysis to examine the language and framing of CSR/ESG reports. Applying corpus linguistics and discourse analysis offers new insights into how banks linguistically construct low-carbon and sustainability narratives, revealing patterns and strategies not captured by traditional quantitative or thematic analyses.This approach in the study can uncover subtle differences in language use, framing, and context between Chinese and American banks, contributing to a deeper understanding of CSR/ESG communication and its potential influence on stakeholders.This also makes such an approach valuable for revealing new dimensions of how low-carbon and sustainability themes are communicated in banking CSR/ESG reports.

2.2 A Review of Studies on Image Construction in Chinese and American Banks

Existing research on the image construction of Chinese banks reveals a multifaceted approach that combines metaphorical, discursive, and visual strategies. Scholars have identified that Chinese banks often employ mechanisms and soil-related metaphors to frame themselves as loyal adherents to national policies, authoritative leaders, and adaptable controllers, emphasizing collective effort (Wang, Hu, & Xu, 2024). In their discursive strategies, Chinese banks highlight historical development, industry rankings, and organizational structures, utilizing numerical data and superlative adjectives to assert their leadership, reflecting a ‘corporate ability’ strategy (Shi & Cheng, 2021). On social media, Chinese banks prioritize entertainment, operational updates, and product promotions, adopting direct and authoritative communication styles. They present themselves as friendly companions, authoritative spokespeople, and innovative competitors, aligning with collectivist cultural values (Sun et al., 2021). In their corporate social responsibility (CSR) discourse, these banks leverage green finance and sustainable development narratives to construct a low-carbon image, consistent with national policy directions (Zhang, 2024). Visual communication in promotional materials frequently employs metaphors, such as depicting banks as ‘strong rescuers’ in news discourse, symbolizing resilience and recovery (Qin, 2023; Yu, 2011).

Previous research on American banks’ image construction also reveals a multidimensional approach, encompassing metaphorical frameworks, discursive strategies, and social media communications to project identities as competitive, service-oriented, and socially responsible entities.American banks rely on different metaphorical frameworks, such as ‘seeding,’ ‘hurt & pain,’ and ‘outcome of game,’ to project identities as competitive winners and proactive stakeholders (Wang et al., 2024). Their discursive strategies focus on employee care, community engagement, and environmental responsibility, using performative verbs to

highlight social contributions—a ‘corporate responsibility’ approach (Shi & Cheng, 2021). On social media, American banks emphasize service-oriented content with indirect, expressive, and involvement-focused communication, reflecting customer-oriented and individualistic values. They position themselves as service providers rather than authoritative figures, though both Chinese and American banks claim the identity of ‘good corporate citizens’ (Sun et al., 2021). In news discourse, they are metaphorically depicted as ‘kings’ and ‘doctors,’ signifying high social status and self-diagnostic capabilities (Qin, 2023). Unlike Chinese banks, they rarely use visual metaphors in promotional materials, relying more on direct service narratives (Yu, 2011).

The comparative research examines the image construction strategies of Chinese and American banks, focusing on metaphorical representation, discursive strategies, social media communication, visual and CSR discourse, and cultural drivers. Chinese banks commonly use mechanistic and soil-based metaphors, such as ‘policy followers’ and ‘resilient rescuers,’ to underscore collectivism and state alignment (Wang et al., 2024; Qin, 2023). In contrast, American banks favor competitive and organic metaphors like ‘game winners’ and ‘self-diagnostic doctors’ to highlight individualism and market leadership (Wang et al., 2024; Qin, 2023). In terms of discursive strategies, Chinese banks prioritize historical achievements and industry dominance, often employing quantitative data to assert their authority (Shi & Cheng, 2021). Meanwhile, American banks emphasize social responsibility and stakeholder care, using qualitative narratives to demonstrate their ethical commitments (Shi & Cheng, 2021). When communicating via social media, Chinese banks adopt direct and authoritative tones with product-focused content, reflecting high-power distance cultures (Sun et al., 2021). Conversely, American banks use indirect and engaging styles with service-oriented content, aligning with low-power distance and individualistic norms (Sun et al., 2021). Regarding visual and CSR discourse, Chinese banks employ visual metaphors in their promotions and link CSR initiatives to national low-carbon policies (Qin, 2023; Zhang, 2024). American banks, on the other hand, minimize the use of visual metaphors and frame CSR as voluntary social contributions (Yu, 2011; Du, 2023). In the context of culture difference, the strategies of Chinese banks reflect collectivist and high-power distance values, while the approaches of American banks align with individualism and low-power distance norms (Shi & Cheng, 2021; Sun et al., 2021). This comparative analysis highlights the influence of cultural drivers on the image construction practices of banks in different national contexts.

Previous investigations into the image construction of Chinese and American banks have primarily focused on metaphorical frameworks, discursive strategies, and cultural drivers. However, they fall short in elucidating the linguistic mechanisms underlying the construction of low-carbon identities. Furthermore, the previous analyses lack systematic, corpus-based approaches to decode the linguistic structure of low-carbon discourses. This study addresses critical gaps by employing corpus linguistics to examine CSR/ESG reports of banks, bridging quantitative and qualitative methods, enhancing cross-cultural comparisons of environmental discourse, and informing practical applications in corporate communication.

2.3 A Review of Studies on Environmental Image Construction in Chinese and American Banks

The existing literature has laid a solid foundation for understanding low-carbon or environmental image construction in banking sectors, yet notable gaps persist, underscoring the value of this study. Little research has systematically analyzed the discursive strategies employed in the CSR reports of Chinese banks, particularly in terms of how high-frequency lexical clusters (e.g., ‘green finance,’ ‘sustainable development’) and metadiscursive elements contribute to shaping a policy-aligned environmental identity (Zhang, 2024; Liu & Zhang, 2024). Some research focuses on the impact of regulatory policies and green finance initiatives on the low-carbon transition and environmental image construction of Chinese banks. It highlights how these factors influence lending decisions and corporate environmental responsibility (CER), particularly in the context of banking competition and transition risks to the financial system. Research on Chinese banks indicates that regulatory policies and green finance initiatives are primary drivers of low-carbon and



environmental image construction. The People's Bank of China's integration of green finance into the Macro-Prudential Assessment framework incentivizes banks to support environmentally responsible enterprises, resulting in lower debt costs for firms with high ESG ratings (Hu, Hong, & Li, 2023). Empirically, banks increasingly incorporate firms' carbon emissions into their lending decisions: higher emissions restrict access to new loans and shorten loan terms, particularly for state-owned enterprises and those subject to strict environmental enforcement (Ren et al., 2023). Low-carbon economic policies also enhance corporate environmental responsibility (CER) by alleviating financial constraints, while bank agglomeration further promotes green practices—especially among private enterprises in market-oriented regions (Liu et al., 2025; Shen et al., 2023). However, the effects of banking competition are mixed: it may increase carbon emissions through fixed asset investments, yet it can also stimulate green innovation and manufacturing servitization when aligned with strong government support for green initiatives (Xiang et al., 2023; Anwar et al., 2024). Transition risks to the financial system are apparent, as stricter regulations elevate default rates among high-polluting firms. Joint equity commercial banks with superior corporate governance manage these risks more effectively than state-owned banks (Punzi, Wu, & Huang, 2021). Although the low-carbon city pilot policy reduces air pollution, it has been shown to hinder regional banking growth by increasing enterprise credit risk (Wang, Liao, & Li, 2022).

American banks are increasingly adopting environmental management strategies, with enhanced environmental performance linked to a reduction in financial tail risk. This relationship is influenced by factors such as corporate governance, CSR committees, and institutional ownership, and is particularly robust during crises (Chaudhry, Ahmed, & Saeed, 2021). The banking sector plays a critical role in addressing climate change through green banking practices, which include eco-friendly programs and green financial products. However, effective regulation and incentives are essential for their adoption (Saif-Alyousfi & Alshammari, 2025). In the US, green finance products encompass green securities, investments, and credit, shaped by environmental policies, risk considerations, and banking regulations (Tenakwah et al., 2021). Major banks, such as Bank of America, have shifted towards sustainability since the early 2000s, driven by public pressure and activism, including a reduction in fossil fuel investments (Elmore, 2021). Voluntary climate disclosures are employed to project a positive institutional image; however, the quality of these disclosures varies with bank size and profitability (Lamarque, Ziane, & Caby, 2020). While American banks have integrated carbon risk pricing into syndicated loans following the Paris Agreement, the risk premium remains low, and 'green' banks do not show significant differences in this approach (Ehlers, de Greiff, & Packer, 2021). Strategies such as online banking, green loans, and sustainable infrastructure aim to enhance green brand images, despite facing implementation challenges (Bansal et al., 2024).

The summary table below outlines key contrasts between Chinese and American banks in low-carbon or environmental image construction. Chinese banks demonstrate a strong regulatory-driven approach, with green finance policies integrated into risk assessment frameworks, leading to lower debt costs for environmentally compliant firms and strict carbon emission-linked lending decisions. Bank agglomeration and low-carbon policies enhance corporate environmental responsibility (CER), though banking competition yields mixed emission effects. Transition risks from stricter regulations are managed differently based on bank ownership. In contrast, American banks adopt green practices driven by sustainability goals and voluntary disclosures, with limited carbon risk pricing in loans. CER is linked to corporate governance and institutional ownership, and banks prioritize green branding and climate disclosure over regulatory mandates.

Table 2: A Comparative Analysis of Low-carbon Image Construction Between Chinese and American Banks

Aspect	Chinese Banks	American Banks
Green Finance Policies	Strong regulatory push, integration into risk assessment, lower debt costs for green firms (Saif-Alyousfi & Alshammari, 2025)	Banks adopt green-banking practices, driven by sustainability goals and regulatory incentives (Saif-Alyousfi & Alshammari, 2025; Tenakwah et al., 2021).
Lending & Carbon Emissions	Higher emissions reduce loan access and terms; credit rating moderates effect (Tenakwah et al., 2021; Bansal et al., 2024)	Some carbon risk is priced into loans post-Paris Agreement, but risk premium is low and limited in scope (Ehlers et al., 2021).
Corporate Environmental Responsibility (CER)	Enhanced by low-carbon policies and bank agglomeration (Elmore, 2021; Lamarque et al., 2020)	Environmental strategies reduce financial risk; influenced by governance, CSR, and institutional ownership (Chaudhry et al., 2021).
Banking Competition	Mixed effects: can increase emissions via investment, but also promote green innovation (Ehlers et al., 2021; Ahmed et al., 2025)	Not a major focus; more emphasis on voluntary climate disclosure and green brand image (Lamarque et al., 2020; Bansal et al., 2024).
Transition Risks	Stricter regulations increase default risk for polluters; governance matters (Chaudhry et al., 2021)	Banks face challenges in implementing green strategies but are pivotal in climate transition (Saif-Alyousfi & Alshammari, 2025; Bansal et al., 2024)

2.4 Conclusion and Critical Review

The current study makes significant contributions to the literature by analyzing high-frequency low-carbon terms, low-carbon word clusters, and employing keyword-in-context (KWIC) analysis. It reveals the discursive construction of low-carbon narratives, demonstrating how Chinese and American banks shape their environmental images through the strategic use of specific terminology. For example, KWIC analysis reveals whether the term “low carbon” is portrayed in Chinese policy documents as a regulatory requirement, such as “mandatory emission reductions,” while in US bank disclosures it is presented as a voluntary commitment, as seen in “sustainability initiatives” (Hu et al., 2023; Saif-Alyousfi & Alshammari, 2025). Additionally, this study bridges the gap between quantitative and qualitative research, which typically focuses on the numerical effects of policies or financial results, such as loan terms and default rates, but often neglects the qualitative intricacies of how “low carbon” is defined and put into action. By employing discourse analysis, this study decodes the semantic frameworks and ideological disparities in environmental discourses, such as the contrast between the ‘state-driven sustainability’ prioritized by Chinese banks and the ‘market-led green innovation’ emphasized by their US counterparts (Elmore, 2021; Xiang et al., 2023). Additionally, the contextual dynamics of policy implementation are uncovered through KWIC and word cluster analysis, revealing that in Chinese banks, ‘carbon risk’ frequently co-occurs with terms like ‘regulatory compliance’ and ‘state-owned enterprises,’ whereas in US banks, it aligns more with ‘voluntary disclosure’.



sure’ and ‘brand reputation’ (Punzi et al., 2021; Lamarque et al., 2020). Lastly, by systematically comparing discursive patterns, this research offers a deeper understanding of how cultural and institutional differences influence the construction of environmental images. It highlights that Chinese discourses often emphasize collective responsibility and state-led initiatives, while US discourses focus on individual enterprise agency and market mechanisms (Liu et al., 2025; Chaudhry et al., 2021). These insights are vital for informing global sustainability strategies, illustrating the distinct linguistic and ideological constructions of ‘low carbon’ across varying contexts.

3 Methodology and Research Design

3.1 Research Data

This study relies on the 2023 Fortune 500 Companies ranking. In order to reduce industry-heterogeneous interference, samples were selected from similar listed banks, including the top four banks in China and the United States, to serve as a representation of the banks in both countries. This was done to minimize the large bias in text discourse that can arise from differences in report types. In the process of selecting a corpus and excluding the diversified Banks of America corpus of the CSR column among the top banks in the two countries, the final corpus came from the English websites of the four major banks of China including Industrial and Commercial Bank of China, China Construction Bank, Agricultural Bank of China, and Bank of China, and four American banks including JPMorgan Chase, Citibank, Wells Fargo, U.S. Bank. During the data screening process, it was discovered that the banks in the two countries have different names for their corporate social responsibility reports. For instance, in the English version of the report, the majority of the four major banks in China are titled “Corporate Social Responsibility Report” (CSR Report). In comparison, the four banks in the United States tend to name the unified statement Environmental, Social, and Governance (ESG Report), which will be referred to as the report from now on.

Among them, the existing reports of Wells Fargo and Citibank can be traced back to 2019. In accordance with the principles of integrity and time-series alignment, the starting year of the reports is kept consistent to enhance the timeliness and representativeness of the analysis. The reports of 2019-2022 including CSR or ESG reports are downloaded from the official websites of these major banks mentioned above. Finally, we collected 33 PDF documents and the detailed data are listed in table 3.

Table 3: The Corpus Data

Corpus	The report source	Number of reports	Number of articles	Number of characters
The Chinese bank corpus	Industrial and Commercial Bank of China (ICBC)	4	16	518079
	China Construction Bank (CCB)	4		
	Agricultural Bank of China (ABC)	4		
	Bank of China (BOA)	4		
The US bank corpus	J.P. Morgan Chase & Co. (JPMorgan Chase)	4	17	401319
	The United States Bank (U.S.Bancorp)	5		
	Citibank (Citigroup)	4		
	Wells Fargo Bank (Wells Fargo)	4		
Total		33	33	919398

3.2 Research Questions

This study employs both quantitative and qualitative methods to investigate the low-carbon image of Chinese and American commercial banks. The research is designed to address the following questions:

What low-carbon image have Chinese and American commercial banks been constructing respectively?

How do Chinese and American commercial banks construct their low-carbon image respectively?

3.3 Research Procedure

To facilitate linguistic analysis, all collected PDF documents are converted into plain text (TXT) format. A text-cleaning procedure is then applied to remove formatting inconsistencies, interfering metadata, and non-textual elements etc., ensuring the accuracy and consistency of the data set.

This study adopts a combined qualitative and quantitative approach including corpus data analysis and discourse analysis and the subsequent steps are outlined in the following figure.

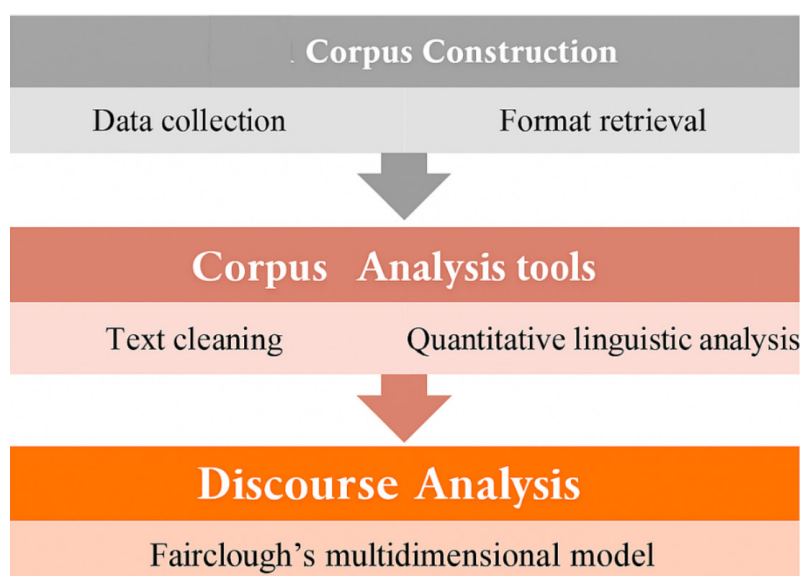


Figure 1. Flow chart of the research procedure

At the textual level, the analysis focuses on linguistic features such as lexical choices, modality, evaluative language, and metaphorical expressions used to frame low-carbon initiatives. This helps reveal how banks position themselves in relation to environmental responsibility.

At the discursive practice level, attention is paid to how these texts are produced, distributed, and consumed. This includes examining intertextuality, the citation of national policies or international agreements (e.g., the Paris Agreement), and how banks adapt their discourse for different stakeholders such as governments, investors, and the public.

At the social practice level, the study interprets how CSR and ESG discourse reflects and reinforces broader socio-political contexts, including China's dual-carbon goals and global climate governance. This dimension explores how sustainability discourse in banking is shaped by and contributes to institutional ideologies and power relations.

By applying this model, the study provides a deeper understanding of the discursive strategies used by Chinese and American banks in constructing a low-carbon corporate image. It highlights both contextual similarities and culturally rooted differences in sustainability communication.

3.4 Research Instrument

This study employs AntConc 4.2.4, a corpus analysis software, as the research tool to systematically examine the linguistic strategies underpinning low-carbon image construction in bank reports. Specifically, the tool's word frequency statistics function is used to generate a high frequency word list table, from which the top 100 high-frequency content words related to "low carbon" are extracted for focused analysis. The collocation network function is then applied to map co-occurring word patterns, revealing semantic associations and discursive frameworks in CSR/ESG reports. Additionally, the concordance line (index line) function is utilized to analyze high-frequency "low-carbon" related terms, enabling a detailed exploration of contextual usage and discourse strategies. These tools collectively facilitate a rigorous investigation into how Chinese and American banks linguistically construct their low-carbon identities.

4 Findings and Discussions

4.1.1 The construction of low-carbon images by Chinese commercial banks

As illustrated in the table, the most frequently occurring words, such as "green" (Freq_Tar = 2891, Freq_Ref = 294), "development" (Freq_Tar = 2782, Freq_Ref = 820), "carbon" (Freq_Tar = 996, Freq_Ref = 447), and "energy" (Freq_Tar = 921, Freq_Ref = 719), highlight the essential components in the construction of a low-carbon image for Chinese banks.

Table 4. Top 20 low-carbon related words by Chinese banks

High-frequency Low-carbon Words of Chinese Bank			
Rank	Word	Freq_Tar	Freq_Ref
1	green	2891	294
2	development	2782	820
3	protection	1136	92
4	carbon	996	447
5	energy	921	719
6	environmental	854	971
7	inclusive	727	280
8	esg	715	1688
9	industry	710	308
10	construction	664	72
11	operation	661	50
12	industries	557	66
13	industrial	550	32
14	resources	445	387
15	climate	418	1284
16	water	411	179
17	sustainable	401	1251
18	transformation	395	46
19	emissions	382	593
20	environment	360	194

Note: "Freq_Tar" refers to the frequency of occurrence of the high-frequency word in the reports of Chinese banks, while "Freq_Ref" represents the frequency of occurrence in the reports of American banks.

The high frequency of the term “green” in Chinese banks’ discourse signifies that their low-carbon efforts are closely associated with the concept of greenness. This implies a strategic alignment of their initiatives with green development objectives. The frequent occurrence of the word “protection” indicates a strong emphasis on safeguarding the environment and resources, which is central to their low-carbon agenda. The prominence of “development” underscores the banks’ focus on integrating low-carbon practices with sustainable development goals, presenting low-carbon actions as drivers rather than obstacles to long-term growth for both the banks and the economy. Additionally, the recurring terms “construction” and “operation” reflect the banks’ active engagement in building and implementing low-carbon infrastructure and operational frameworks. In general, these high - frequency words collectively demonstrate that Chinese banks construct their low-carbon image by integrating low-carbon efforts into the narratives of green development, environmental protection, and practical implementation. Through the frequent use of such words in their ESG/ CSR reports, they aim to showcase their commitment to low-carbon development and establish a positive low-carbon corporate image.

To gain a more lucid insight into the specific low-carbon image crafted, a word cloud clustering of low-carbon associated terms was generated, as presented in the figure 2.

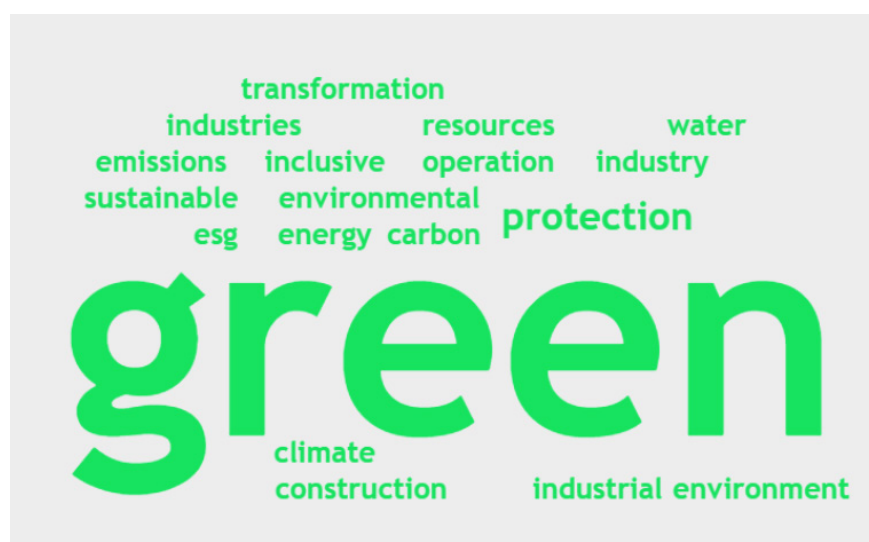


Figure 2. Low-carbon related word cloud by Chinese banks

In the word cloud, the term “green” is notably larger than other related terms such as “protection,” “carbon,” “energy,” “environmental,” “sustainable,” “ESG,” and “climate.” This visual significance of “green” suggests that Chinese commercial banks base their low-carbon image construction on the central idea of “green development.” The grouping of words like “environmental protection,” “carbon,” and “energy” around “green” indicates that banks are integrating low-carbon initiatives into a wider green ecosystem, highlighting their commitment to environmental stewardship, carbon management, and sustainable energy use.

Additionally, other clustered terms such as “transformation,” “industries,” “construction,” and “industrial environment” imply that Chinese banks associate low-carbon efforts with industrial transformation and the development of sustainable infrastructure. Words like “inclusive,” “operation,” and “resources” further suggest that the banks’ low-carbon practices are framed as all-encompassing, addressing inclusive growth, operational efficiency, and resource optimization.

After understanding that Chinese commercial banks create a low-carbon image characterized by a strong “green” narrative, we further generated a word cloud for the term “green”(figure 2) to delve deeper into the specific constructions and developments that Chinese banks focus on in relation to the “green” narrative.



Figure 3. The word cloud of “green” by Chinese banks

The word cloud prominently features “green finance” in large, surrounded by a dense network of associated terms such as “green credit”, “green bonds”, “green development”, and “green loans”. This clustering vividly illustrates that Chinese commercial banks anchor their low-carbon image construction within the framework of green financial mechanisms. The centrality of “green finance” signals that banks position themselves as key enablers of low-carbon transitions, leveraging financial instruments to drive sustainable change. Terms like “green credit” and “green bonds” highlight the banks’ role in channeling capital toward low-carbon projects, showcasing their commitment to financing low-carbon initiatives. “Green development” and “green industries” further extend this narrative, indicating that banks support industries in transitioning to low-carbon sectors to achieve China’s dual - carbon goals. Additionally, collocates such as “green procurement”, “green building”, and “green operations” reflect a holistic approach. They demonstrate that Chinese banks embed low-carbon considerations not only in financial products but also in internal processes and service offerings, from sourcing sustainable materials to designing eco - friendly workplaces.

In essence, the cluster of “green”- related terms in the word cloud reveals that Chinese commercial banks construct a low-carbon image centered on green finance as a catalyst. By emphasizing financial innovation, industrial transformation, and operational integration, they project an identity of proactive institutions driving the transition to a low-carbon economy, merging financial responsibility with environmental stewardship. This linguistic and conceptual clustering underscores their strategic communication of a comprehensive, finance - led low-carbon agenda in ESG/CSR reports.

4.1.2 The low-carbon image construction of American commercial banks and its comparison with Chinese commercial banks

To explore the low-carbon images constructed by American banks, this study extracts high - frequency low-carbon - related words from their ESG/CSR reports, with the findings presented in the following table.

Table 5. Top 20 low-carbon related words by Chinese banks

Rank	Word	Freq_Tar	Freq_Ref
1	climate	1284	418
2	sustainable	1251	401
3	sustainability	1028	148
4	environmental	971	854
5	development	820	2782
6	energy	719	921
7	impact	701	189

8	commitment	653	50
9	emissions	593	382
10	opportunities	471	106
11	carbon	477	996
12	change	459	135
13	low	420	313
14	resources	387	445
15	zero	374	35
16	framework	370	100
17	renewable	313	35
18	industry	308	710
19	green	294	2891
20	inclusive	280	727

As shown in the table, “sustainable” (Freq_Tar = 1251; Freq_Ref = 401), “climate” (Freq_Tar = 1284; Freq_Ref = 418), and “sustainability” (Freq_Tar = 1028; Freq_Ref = 148) are among the top - ranked high - frequency terms. The high prevalence of “sustainable” and “sustainability” indicates that American banks tend to embed their low-carbon initiatives within a comprehensive sustainability framework. This linguistic choice reflects their attempt to project an image of aligning with long - term environmental stewardship and ethical imperatives. The prominent frequency of “climate” suggests that these banks frame their low-carbon efforts as integral responses to climate change, positioning themselves as active participants in global climate action.

Other significant terms include “environmental” (Freq_Tar = 971; Freq_Ref = 854), “energy” (Freq_Tar = 719; Freq_Ref = 921), and “emissions” (Freq_Tar = 593; Freq_Ref = 382). The relatively high occurrence of “environmental” reinforces a broad eco - conscious narrative, while “energy” highlights a focus on energy - related low-carbon actions, such as the promotion of renewable energy. The notable frequency of “emissions” points to an explicit emphasis on carbon emissions reduction, a fundamental aspect of low-carbon strategies, which helps shape an image of banks taking concrete steps in carbon management.

Moreover, words like “commitment” (Freq_Tar = 653; Freq_Ref = 50) and “opportunities” (Freq_Tar = 471; Freq_Ref = 106) reveal a strategic communication approach. “Commitment” is used to showcase the banks’ dedication to low-carbon goals, fostering an image of responsibility and accountability. “Opportunities”, on the other hand, frames low-carbon transitions as avenues for innovation and business growth, portraying the banks as forward - looking entities that can capitalize on environmental trends.

In summary, through the high - frequency words in their ESG/CSR reports, American commercial banks construct a low-carbon image characterized by integration into sustainability and climate action narratives, emphasis on environmental responsibility, tangible efforts in emissions reduction, and a dual focus on ethical commitment and business opportunity. These linguistic patterns in their reports serve to project an identity of proactive, responsible, and innovative players in the low-carbon economic landscape.

To gain a clearer understanding of the specific low-carbon image they construct, a cluster of low - carbon - associated terms was generated, as shown in the following figure.

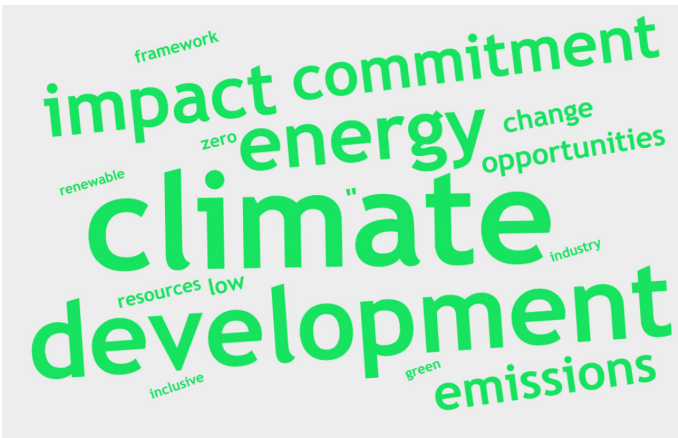


Figure 4. Low-carbon related word cloud by American banks

In this word cloud, “climate” and “development” are prominently displayed in large fonts, surrounded by terms like “energy”, “impact”, “commitment”, “opportunities”, and “emissions”. The central position of “climate” and “development” indicates that American commercial banks anchor their low-carbon image construction around the core concept of climate - centered sustainable development. The clustering of “energy” reflects a focus on energy - related low-carbon initiatives, such as the promotion of renewable energy (implied by the term “renewable” in the cloud).

“Commitment” suggests that these banks aim to project an image of being dedicated to low-carbon goals, while “opportunities” implies that they frame low-carbon transitions as avenues for business growth and innovation. “Emissions” highlights an emphasis on addressing carbon emissions, a fundamental aspect of low-carbon strategies. Terms like “framework” and “change” indicate that American banks may be positioning their low-carbon efforts within structured plans and as part of broader transformative processes.

Collectively, this word cloud reveals that American commercial banks construct a low-carbon image characterized by a strong focus on climate - related sustainable development. They communicate their engagement in energy transition, emission reduction, and the pursuit of business opportunities within the low-carbon economy, while also showcasing their commitment to climate - related responsibilities. This linguistic clustering in the ESG/CSR reports helps shape an image of American banks as proactive participants in global climate action and sustainable development.

It was found that “sustainable” emerged as the most frequently occurring term, signifying its centrality as a core concept in shaping the low-carbon narrative of American banks. To further unpack its significance, AntConc 4.2.4 was utilized to generate a collocate cluster (figure) for “sustainable”, with the resulting word cloud presented in the attached image.



Figure 5. Word cloud of “sustainable” by American banks

The word cloud prominently features “sustainable development”, “sustainable finance”, and “sustainable operations” in large, bold fonts, surrounded by a dense network of associated phrases such as “sustainable progress”, “sustainable investing”, and “sustainable business”. This clustering vividly illustrates that American commercial banks anchor their low-carbon image construction within a comprehensive sustainability - centered framework. The prominence of “sustainable development” signals a strategic emphasis on long - term, holistic growth that integrates environmental stewardship as a fundamental pillar.

Terms like “sustainable finance” and related expressions (e.g., “sustainable investing”, “sustainable bonds”) highlight the banks’ role in leveraging financial mechanisms to drive low-carbon transitions. By positioning themselves as facilitators of sustainable financial flows, these banks project an image of institutions that balance profit - seeking with environmental responsibility. “Sustainable operations” and “sustainable business” extend this narrative to internal processes, demonstrating a commitment to embedding low-carbon principles in day - to - day activities and overarching corporate strategies.

Additionally, collocates such as “sustainable cities”, “sustainable agriculture”, and “sustainable transport” reflect a broad, ecosystem - level perspective. They indicate that American banks frame their low-carbon efforts as contributing to large - scale, multi - sectoral sustainability goals, transcending the boundaries of mere internal operations.

In essence, the cluster of “sustainable” - related terms in the word cloud reveals that American commercial banks construct a low-carbon image centered on sustainability as a core ethos. By emphasizing sustainable development, finance, and operations, they project an identity of proactive institutions driving systemic change toward a low-carbon economy. This linguistic and conceptual clustering in ESG/CSR reports underscores their strategic communication of a comprehensive, sustainability - led low-carbon agenda, positioning themselves as key players in the global transition to net - zero while aligning with broader societal sustainability aspirations.

4.2 Three-Dimensional Model of construction of low-carbon image by Chinese and American banks

The corpus consists of 33 ESG/CSR reports published between 2019 and 2022 by four leading Chinese and American commercial banks.

Every word tends to occur in clusters with other words, providing valuable insights for collocation, association and semantic resonance studies (Yang, 2002). From a linguistic perspective, the study focuses on the collocations, semantic prosody, and evaluative stance expressed in low-carbon-related discourse. Drawing on corpus tools, we examined the keywords “carbon” (Chinese banks) and “net-zero” (American banks) using concordance lines to identify recurring patterns, such as co-occurring words and collocational clusters. These clusters, referred to as “co-text” (Li & Pu, 2001), reveal underlying semantic associations and emotional coloring (Qian, 2010), offering insights into how banks signal commitment or responsibility in environmental discourse.

4.2.1 Textual Analysis

Words often appear in clusters, offering insights for studying collocations, associations, and semantic resonances (Yang, 2002). Using corpus tools, we focused on low-carbon related collocations and clusters in the reports of Chinese and American banks. For Chinese banks, we examined the keyword “carbon,” and for American banks, the keyword “net-zero.” By analyzing concordance lines, we identified recurring patterns such as co-occurring words and collocational clusters. These clusters, or “co-text” (Li & Pu, 2001), reveal underlying semantic associations and emotional coloring (Qian, 2010), shedding light on how banks express their commitment to environmental responsibility.

The high-frequency clusters in the Chinese bank corpus are dominated by the term “green,” along with words like “environmental,” “protection,” “transformation,” “resources,” and “industry.” This pattern reflects a discourse closely aligned with national policies and collective responsibility. Chinese banks tend to construct their low-carbon image through policy-driven and collectively framed language, using terms such as “green solution.” The prominence of keywords like “green,” “protection,” “transformation,” and “industry” indicates a systematic approach where environmental concerns are integrated into narratives of national development and industrial change.

In contrast, American banks emphasize terms like “sustainable development” and “sustainable finance.” The keyword “net-zero” is central to their discourse, reflecting a focus on achieving a balance between carbon emissions and removals. This difference suggests that while Chinese banks highlight green initiatives within a national policy framework, American banks stress sustainable development and specific carbon neutrality goals.

In contrast, American banks emphasize terms like “sustainable development” and “sustainable finance.” The keyword “net-zero” is central to their discourse, reflecting a focus on achieving a balance between carbon emissions and removals. This difference suggests that while Chinese banks highlight green initiatives within a national policy framework, American banks stress sustainable development and specific carbon neutrality goals.

4.2.2 Discursive Practice

At the level of discursive practice, differences between Chinese and American banks become evident in how low-carbon discourse is produced, distributed, and consumed. These processes reflect varying institutional pressures, audience expectations, and sociopolitical contexts.

Chinese banks produce CSR and ESG reports in response to government policies and national strategic plans such as the “Dual Carbon” goals. Their language is intertextual with state documents and policy frameworks, positioning them as responsible actors aligned with national objectives. For example, they emphasize aligning national policies with economic development to enhance public awareness of sustainability. Internally, they integrate low-carbon awareness into daily operations, and socially, they support green projects that contribute to both economic growth and societal well-being. These practices reflect a top-down communicative logic that mirrors state priorities, emphasizing institutional compliance and collective responsibility.

American banks generate reports primarily in response to pressures from the public, investors, and international policy frameworks, resulting in a market-oriented and brand-driven discourse. Their intended audience includes investors, ESG rating agencies, and the general public. American banks use lexical choices like “challenges,” “COVID-19,” and “rise to” to highlight their proactive stance during global disruptions. They also employ international frameworks such as the SDGs and the Paris Agreement to build symbolic authority, enhancing their legitimacy and corporate image. Furthermore, they link sustainability to profitability, framing ESG engagement as a win-win proposition that combines financial performance with social and environmental values.

4.2.3 Social Practice

The differences in discursive practices between Chinese and American banks reflect their distinct socio-institutional environments.

In China, low-carbon discourse is integrated within a top-down governance framework where banks actively participate in and support national strategic priorities. Their ESG language closely aligns with initiatives such as the “Dual Carbon” goals, reflecting a commitment to sustainability that is deeply rooted in national development strategies. This discourse emphasizes coordination with government objectives and policies, highlighting the banks’ role in promoting sustainable economic growth and environmental protec-

tion. It underscores the banks' proactive engagement in achieving national targets, demonstrating their responsibility and dedication to the holistic advancement of the economy, society, and environment. This approach reflects a comprehensive and collaborative effort to achieve sustainable development, where banks contribute to and align with the nation's strategic goals.

In contrast, American banks operate within a liberal market economy where ESG discourse is shaped by the expectations of investors, ESG rating agencies, international organizations, and a socially conscious public. Their language adopts a corporate, forward-looking tone, often invoking global governance frameworks such as the UN Sustainable Development Goals (SDGs) and the Paris Climate Agreement. Unlike the policy-aligned narrative of Chinese banks, American discourse focuses on strategic branding, profitability, and stakeholder engagement. Sustainability is both a moral imperative and a competitive advantage, reflecting the dominance of neoliberal ideology where corporations are seen as autonomous agents responsible for self-regulation and innovation in addressing climate change.

In summary, the analysis of Chinese and American banks' low-carbon image construction through Fairclough's Three-Dimensional Model reveals how textual and discursive choices are deeply rooted in each country's governance structures, economic orientations, and sociopolitical contexts.

5 Conclusion

This study has provided a comprehensively comparative analysis of the low-carbon image construction by Chinese and American commercial banks, examining the language used in their CSR/ESG reports from 2019 to 2022 through corpus linguistics and discourse analysis. The research addressed two fundamental questions: (1) What kind of low-carbon image have Chinese and American commercial banks respectively constructed? (2) How do these banks construct their low-carbon image? The corpus tool, AntConc 4.2.4 was employed to examine the linguistic features in the reports and three-dimensional discourse analysis model is deployed to examine the strategies of constructing low-carbon image. The findings reveal that Chinese banks have constructed a low-carbon image centered on green development and policy alignment, emphasizing their role in supporting national goals such as China's "dual-carbon" targets. They achieve this by integrating green narratives with national development strategies and utilizing green financial instruments like "green finance" and "green bonds." In contrast, American banks have built a low-carbon image focused on global climate governance and sustainability, highlighting their participation in international climate initiatives and commitments to sustainable development. They employ a market-oriented approach, leveraging global frameworks such as the Paris Agreement and SDGs to enhance their corporate image and engage stakeholders. Although, both banks effectively communicate their dedication to low-carbon development, their strategies reflect their distinct socio-political and economic contexts.

This research offers a substantial theoretical advancement within the interdisciplinary realm of corporate environmental discourse by employing corpus-based methodologies to investigate the cross-cultural variances in sustainability communication. This approach enhances our comprehension of how environmental identities are linguistically constructed, unveiling subtle linguistic trends that traditional analyses often overlook, thereby enriching the dialogue surrounding CSR and ESG communication. On a practical level, the insights gained equip banks with strategies to refine their low-carbon communication efforts in accordance with global climate governance frameworks, highlighting the distinctive ways in which Chinese and American banks present their sustainability narratives. This understanding is essential for effectively communicating sustainability pledges and positively influencing stakeholder perceptions. Future inquiries could expand their focus by incorporating additional countries and banking sectors. Furthermore, it is crucial to explore how these discursive strategies impact stakeholder behavior and investment decisions.



Funding

This study is a phased achievement of the project titled “Research on the Construction of Chinese Enterprises’ Overseas Social Responsibility Identity Discourse in the Context of ‘Cultural Confidence’,” which is funded by Guangdong Planning Office of Philosophy and Social Science (Grant No.: GD24XWY08).

References

- Anwar, S., Chen, X., & Xu, H. (2024). Bank competition, government interest in green initiatives and carbon emissions reduction: An empirical analysis using city-level data from China. *The North American Journal of Economics and Finance*, 70, <https://doi.org/10.1016/j.najef.2024.102144>
- Chen, T., Ho, J., & Wu, J. (2021). Are corporate social responsibility reports informative? The evidence from textual analysis of banks in China presented in the article “China Finance Review International” provides valuable insights into the financial landscape of the country. <https://doi.org/10.1108/cfri-04-2021-0081>
- Chaudhry, S., Ahmed, R., & Saeed, A. (2021). Carbon neutrality: The role of banks in optimal environmental management strategies. *Journal of Environmental Management*, 299, 113545. <https://doi.org/10.1016/j.jenvman.2021.113545>
- Cong, M. (2023). The linkage mechanism of carbon finance promoting low-carbon development from the perspective of dual carbon policy. *Highlights in Business, Economics and Management*, 2023-01-20.
- Ehlers, T., de Greiff, K., & Packer, F. (2021). The pricing of carbon risk in syndicated loans: Which risks are priced and why? SRPN: Carbon Reduction (Topic). <https://doi.org/10.2139/ssrn.3857716>
- Elmore, B. (2021). The environmental history of an American bank. *Environmental History*, 27, 113–139. <https://doi.org/10.1086/717439>
- Ervits, I. (2021). CSR reporting in China’s private and state-owned enterprises: A mixed methods comparative analysis. *Asian Business & Management*, 22, 55–83. <https://doi.org/10.1057/s41291-021-00147-1>
- Feng, J., Goodell, J., & Shen, D. (2021). ESG rating and stock price crash risk: Evidence from China. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2021.102476>
- Fu, X., Meng, W., Wu, J., & Wang, D. (2023). Antecedent configurations of ESG disclosure: Evidence from the banking sector in China. *Sustainability*. <https://doi.org/10.3390/su151713234>
- Fu, Y., Guo, J., & Zhao, X. (2025). Green finance, technological innovation, and low-carbon development in logistics industry: mechanisms and threshold effects. *Environment, Development and Sustainability*, 2025-01-06.
- Guo, X. (2025). Research on the risk management of commercial bank green bonds under the “dual carbon” goals. *Frontiers in Business, Economics and Management*, 2025-01-21.
- Hu, H., Hong, Z., & Li, W. (2023). Green finance policy, ESG rating, and cost of debt—Evidence from China. *International Review of Financial Analysis*, 89. <https://doi.org/10.1016/j.irfa.2023.103051>
- Hua, H., & Zhang, Y. (2025). Carbon financial system construction under the background of dual-carbon targets: current situation, problems and suggestions. 2025-02-19.
- Lamarque, E., Ziane, Y., & Caby, J. (2020). The determinants of voluntary climate change disclosure commitment and quality in the banking industry. *Technological Forecasting and Social Change*, 161, 120282. <https://doi.org/10.1016/j.techfore.2020.120282>
- Li, N. (2023) conducted a study on the development of green transformation and development of commercial banks in the “double carbon” background. *Research in Economics and Management*, 2023-08-19.
- Li, W. Z., & Pu, J. Z. (2001). The application of corpus indexing in foreign language teaching. *Jiefangjun Waiguoyu Xueyuan Xuebao*, 24(2), 20–25.
- Liu, H., Ibrahim, H., & Song, M. (2025). Impact of low-carbon economic policies on the corporate environmental responsibility model in China. *PLOS ONE*, 20(1), e0314589. <https://doi.org/10.1371/journal.pone.0314589>

- Liu, X.-Q., & Zhang, Y.-Y.(2024) conducted a study on the construction of environmental image of four Chinese banks using metadiscourse analysis.*Advances in Social Sciences Research Journal*, 10(29).
- Punzi, M. T., Wu, Y., & Huang, B. (2021). Do banks price environmental transition risks?The study conducted in China provides valuable insights into the effects of corporate governance on firm performance. The findings suggest that there is a significant relationship between corporate governance mechanisms and financial outcomes. This research adds to the existing literature by highlighting the importance of effective governance structures in enhancing firm value and overall performance. The results of this quasi-natural experiment contribute to our understanding of how corporate governance practices can impact the financial success of companies.<https://doi.org/10.1016/j.jcorpfin.2021.101983>
- Qian, Y. F. (2010). Corpus and critical discourse analysis. *Waiguo Jiaoxue yu Yanjiu*, (3), 198–202.
- Ren, Y., Wu, H., Ding, X., & Tan, W. (2023). Do carbon emissions of firms matter for bank loan decisions?Evidence from China. *International Review of Financial Analysis*, 86, 102556.<https://doi.org/10.1016/j.irfa.2023.102556>
- Saif-Alyousfi, A. Y. H., & Alshammari, T. (2025). Environmental sustainability and climate change: An emerging concern in banking sectors.*Sustainability*, 17(3), 1040. <https://doi.org/10.3390/su17031040>
- Shen, M., Liu, Y., Tan, W., & Pang, Y. (2023). Bank agglomeration and corporate environmental responsibility performance: Evidence from Chinese bank branch data. *International Review of Financial Analysis*, 88, 103021. <https://doi.org/10.1016/j.irfa.2023.103021>
- Shi, X., & Cheng, X. (2021). A corpus-based study of the discursive construction of corporate identities by Chinese and American banks.*Contrastive Pragmatics*, 2(2), 1–23. <https://doi.org/10.1163/26660393-12340008>
- Sun, Y., Wang, G., & Ren, H. (2021). To entertain or to serve: Chinese and US banks' online identity based on a genre analysis of social media.*IEEE Transactions on Professional Communication*, 64(2), 123–137. <https://doi.org/10.1109/TPC.2021.3064395>
- Sun, Y., Liu, J., Chiu, Y., & Ji, L. (2022). Environmental, social, and governance (ESG) and market efficiency of China's commercial banks under market competition. *Environmental Science and Pollution Research International*, 30, 24533–24552. <https://doi.org/10.1007/s11356-022-23742-x>
- Tunio, R., Xiao, J., Mirani, A., Das, G., Laghari, M., & Jamali, R. (2020). The relationship between corporate social responsibility disclosures and financial performance: a mediating role of employee productivity.*Environmental Science and Pollution Research*, 28, 10661–10677. <https://doi.org/10.1007/s11356-020-11247-4>
- Wang, J., Hu, C., & Xu, Y. (2024). Metaphor scenarios, stances, and corporate identities.*Ibérica*, 48, 295–320. <https://doi.org/10.17398/2340-2784.48.295>
- Wang, C., Liao, L., & Li, X. (2022). Can environmental protection and banking development be synergized?An analysis based on the low carbon city pilot in China. *Environmental Science and Pollution Research*, 29(43), 67584–67599.<https://doi.org/10.1007/s11356-022-22725-2>
- Wang, C., Yue, Z., & Wang, Q. (2024). Research on the impact of green finance on industrial structure in Shaanxi Province.*Academic Journal of Business & Management*, 2024-01-01.
- Xiang, Y., Bi, Y., & Cui, H. (2023). The impact and channel effects of banking competition and government intervention on carbon emissions: Evidence from China.*Energy Policy*, 177, 113476. <https://doi.org/10.1016/j.enpol.2023.113476>
- Xu, Z. (2025). Green finance evolution and corporate innovation in China: Paving the way to dual carbon targets.*Advances in Economics and Management Research*, 2025-03-28.
- Yang, D., Wang, L., & Zhao, Z. (2024). Is ESG Reporting Informative to Debtholders?— Evidence from a Quasi-Experiment in China. *Finance Research Letters*.<https://doi.org/10.1016/j.frl.2024.105100>
- Yang, H. Z. (2002).Introduction to corpus linguistic. Shanghai: Shanghai Waiyu Jiaoyu Chubanshe.
- Yu, H. (2011). Visualizing banking and financial products: A comparative study of Chinese and American practices.*Journal of Technical Writing and Communication*, 41(3), 243–260.



- Yuan, X., Li, Z., Shang, L., & Xu, J. (2021). ESG disclosure and corporate financial irregularities – Evidence from Chinese listed firms. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2021.129992>
- Zhang, M., Zhou, Z.-J., Ni, Z., & Kang, T.-K. (2022). Evaluation of the effect of green finance on low-carbon development in the context of "dual carbon". *International Journal of Education and Humanities*, 2022-11-27.
- Zhang, Y. (2024). A corpus-based analysis of low-carbon image construction in Chinese commercial banks. *English Language Teaching and Linguistics Studies*, 6(3), 45–62.
- Zhao, Y., & Zhu, B. (2022). Carbon risk and the cost of bank loans: Evidence from China. *Technological Forecasting and Social Change*, 181, 121741. <https://doi.org/10.1016/j.techfore.2022.121741>
- Zhang, Y.-J., & Du, M. (2025). The synergistic carbon emission reduction advantage of green finance and digital finance. *Environmental Impact Assessment Review*, 2025-03-01.

Publication Ethics and Malpractice Statement

The *Innovation in Applied Linguistics* is an academic journal published by Mason Publish Group. It is necessary to agree upon standards of expected ethical behavior for all parties involved in the act of publishing: the journal editor, the author, the peer reviewer and the publisher. This ethic statement is based on COPE's Best Practice Guidelines for Journal Editors.

Duties of Editor

- The evaluation should be objective and fair, and give clear opinions on whether it meets the standards of public publication.
- Ensure that journals are published on time and on schedule, and prevent plagiarism and repeated publication of papers.
- According to the importance, originality and relevance of the paper to the journal, the author has the right to reject and accept the paper, and allows the author to appeal the review decision.
- The editor shall ensure that all research material they publish conforms to internationally accepted ethical guidelines.

Duties of Authors

- Authors have to make sure that they write completely original works that have not been published elsewhere before.
- The data of the paper is authentic and reliable, and forgery, tampering and plagiarism are prohibited.
- Reference information shall be indicated when quoting opinions.
- The author shall ensure that the published paper is authorized and approved by the coauthor, and there is no signed dispute.
- All authors must contribute to the research of the paper.
- The author shall provide the original data related to his manuscript for the editor's review, and must keep these data.

Duties of Reviewers

In the process of review, the reviewer shall be objective and keep all information of the paper confidential. Reviewers must ensure that the author has identified all data sources used in the study. If invited to review a paper, it shall be completed within the specified time.

Duties of Advisory Board Members

The advisory board is composed of distinguished scholars and professors in the field of Country, Area and Advanced Technology. Members can review submitted manuscripts, identify topics for special issues, or attract new authors and submissions as necessary.

Innovation in Applied Linguistics

Call for Papers

Language innovation lies at the heart of human progress, transforming education, culture, and global connectivity in our interconnected world. Innovation in Applied Linguistics invites researchers, educators, and practitioners worldwide to contribute cutting-edge insights into the evolving challenges, breakthroughs, and future directions of linguistic theory and practice. We welcome original research, empirical studies, and critical reviews that connect language innovation with pedagogy, technology, policy, and real-world applications. Join us in advancing the frontiers of linguistic knowledge and fostering cross-disciplinary collaboration across the dynamic landscape of applied linguistics.

Call for Reviewers

In order to further improve the academic level and quality of the journal, enrich the team of reviewers and implement the anonymous reviewer system, the editorial department is now calling for reviewers. We welcome experts and scholars in the fields of the Innovation in Applied Linguistics to recommend themselves or recommend suitable candidates, so as to jointly promote the progress and development of periodicals. Please contact us (editorialoffice@masonpublish.org) if you are interested.

2025



Mason Publish Group
<https://masonpublish.org/>