

# **Modern Languages, Literatures, and Linguistics**

Vol. 2 No. 02 2023  
[www.jhkpress.com](http://www.jhkpress.com)  
JHK Press, USA



ISSN 2834-5207

# MLLL

Modern Languages, Literatures, and Linguistics (MLLL) publishes genuinely original research articles, remarks and replies and article reviews. As an international peer-reviewed research Journal, MLLL welcomes original research papers focusing on language acquisition, language teacher education, translation and literary criticism. The issues also cover various linguistics topics including phonetics and phonology, lexis, grammar, pragmatics, and discourse. MLLL uses double-blind review: the reviewers' identities remain anonymous to authors and vice versa. Manuscripts submitted to the journal first go through an initial screening by the editorial team. Those that clear the screening are then sent to at least two experts for peer review. Email address for submitting your manuscript: [mlll@jhkpress.com](mailto:mlll@jhkpress.com).

All articles in this issue are open-access, distributed under the terms of the Creative Commons Attribution Licence (CC-BY) 4.0 <https://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.



## Table of Contents

A Study on the Translation of Mao Zedong's Poetic Dynamic Images from the Perspective of Image Schema Theory- 1 -	Haowen LIANG
A ChatGPT-driven Study of Translating and Post-editing Chinese Enterprise Publicity Materials..... - 7 -	Junchao Wang
A Historical Comparative Study on the Annual Reports of Chinese Public Companies..... - 18 -	Chaowang REN; Haolin ZHENG

-

# A Study on the Translation of Mao Zedong's Poetic Dynamic Images from the Perspective of Image Schema Theory

Haowen LIANG<sup>1</sup>

<sup>1</sup>*Guangzhou Xinhua University, Guangzhou, China 874664699@qq.com*

This Article is a part of the author's Visiting Scholar Program in South China University of Technology.

## Abstract

Mao Zedong's poetry is highly valued for its literary and ideological significance, as it is rich in allusions and image that reflect his patriotism and philosophical thinking. Additionally, his poems have also been translated by renowned translators such as Xu Yuanzhong. To better translate Mao Zedong's poetry, this paper focuses on the analysis of the dynamic images with movement connotations in his poetry, and makes use of the image schema theory of cognitive linguistics to compare and analyze the images and corresponding translations of relevant verses, in order to explore the gains and losses of different translations and summarize translation methods for Chinese classical poetry, thus helping Chinese culture to go global.

**Keywords:** poetry translation; image schema theory; Mao Zedong's poetry; dynamic images

## 1. Introduction

Mao Zedong is a politician, and an outstanding poet in China. His poetry, on the one hand, is characterized by patriotism, philosophical thinking, and rich allusions. On the other hand, like Chinese classical poetry, it is rich in image and achieves the effect of "poetry in painting, painting in poetry". With rich image, it expresses specific meanings and special emotions. Among many images, there is one of type that can be called dynamic image, for lack of a better term, and are often seen in his poems. For example, in "Nian Nu Jiao: Kunlun", there is the image of "Your three million white jade dragons in flight/ Freeze the sky with piercing cold", which actually compares the ice and snow on Kunlun Mountain to millions of jade-like flying dragons. At the same time, the word "flying" in the sentence highlights the dynamic aspect, depicting the dangerous and steep natural environment of the Mountain shrouded in snow. Another example is in "Qing Ping Yue - Jiang Gui War", where the phrase "Over River Ting our red flags leap; To Longyan and Shanghang we sweep" vividly describes the rapid advance and the fearlessness of the Red Army. Because English is a static language, the presentation of dynamic images in poetry become a challenge in language translation. If Mao Zedong's poetic image (especially dynamic image and chains of such images) cannot be well understood, readers of other linguacultural groups will not be able to understand Mao Zedong's poetry. Therefore, it is necessary to borrow the image schema theory from cognitive linguistics to decipher the image code in poetry and better translate Mao Zedong's poetry into English culture.

## 2. Literature Review

### 2.1 Cognitive Linguistics Approaches to Chinese Poetry Translation

Currently, research on the translation of Chinese classical poetry from the perspective of cognitive linguistics mainly includes: 1) studying the translation of cultural elements in Chinese classical poetry by combining cognitive linguistics (Liu, 2014; Zhang & Jiang, 2019). Among them, Li (2014:128-130) and Zhang & Jiang (2019:29-32) believe that translators should fully understand the cultural backgrounds of the original author, the translator, and the target readers in order to translate the relevant images. 2) Analyzing the translation of image schema and metaphors in Chinese classical poetry from the perspective of cognitive linguistics (Chi, 2011; Deng & Zhong, 2012; Huang, 2019). Among them, Chi (2011:75-77) summarizes three situations that translators and target language readers face when understanding symbolic words: firstly, the image schema of the source language and the target language are consistent; secondly, the image schema of the source language and the target language are different; thirdly, the image schema of the source language is missing in the target language. Corresponding translation methods are summarized for each situation. Deng & Zhong (2012:77-82) point out that the image in Chinese classical poetry usually exists in specific scene descriptions or spatial relationships, which are closely related to image schema theory, and can be used to guide the translation of classical poetry. Huang (2019:105-112), combining the perspective of cognitive metaphor theory, summarizes six types of metaphors in Su Shi's works, and points out that translators should flexibly use both alienation and domestication strategies based on the objective experience of the text and the subjective cognition of the translator when translating. Overall, current research with the cognitive linguistics approaches does not subdivide the image schema in classical poetry, and research on dynamic image schema is scarce.

### 2.2 Research on the Translation of Mao Zedong's Poetry

Research on the translation and dissemination of Mao Zedong's poetry falls into the following two aspects: the study of specific language phenomena, and the analysis of the translation and dissemination of Mao Zedong's poetry from a cross-cultural perspective. Specific language phenomena can be further subdivided into words and sentences (Long, Liu & Liu, 2013; Wang, 2007; Yu, 2016). Long, et al. (2013) uses the blended space theory framework to study the translation of cultural images in Mao Zedong's poetry, proposing methods such as literal translation, free translation, free translation with annotations, and image replacement to convey cultural images. Wang (2007) analyzes the translation of several lines of Mao Zedong's poetry containing the word "fly" by exploring the deep meaning in the original poem and combining different interpretations by other famous scholars. He compared multiple translations by foreign publishers and Xu Yuanchong, and ultimately pointed out that when translating poetry, translators must first consider the logical relationship implied in the original sentence, followed by the language structure of the original poem as the second dimension to be considered.

Besides, both the logical relationship and the language structure in the original text should serve the overall meaning structure. Many scholars have a review and analysis of the translation and dissemination of Mao Zedong's poetry from the cross-cultural perspective (e.g., Li, Li & Li, 2008; Yu, 2016; Qi, 2020; Li et al., 2008) classify the rich allusions in Mao Zedong's poetry into three categories: legendary fables and stories, characters, and sourced words and phrases. After comparing eight translations by famous scholars, they summarized the translation methods for different categories. Yu (2016) analyzes and studied the translation of five culture-loaded words in Mao Zedong's poetry by Xu Yuanchong and Barnstone respectively, based on the classification of culture-loaded words by the translation master Nida and the purpose theory of the German functionalist school. Qi (2020) systematically investigates and studied the cross-cultural dissemination of Mao Zedong's poetry since the founding of New China based on the principles of cross-cultural communication, and sorted through relevant historical materials. It is clear that current research on the translation of Mao Zedong's poetry mainly focuses on traditional translation studies and the paradigm of cross-cultural communication. Although some studies have noticed the rich image and lines containing dynamic processes in Mao Zedong's poetry, researchers have not combined other theoretical paradigms (such as cognitive linguistics) for research. Therefore, it is promising to use the paradigm of cognitive linguistics to analyze the dynamic image in Mao Zedong's poetry.

### *2.3 Overview of Image Schema Theory*

The Image Schema Theory of cognitive linguistics is related to the image found in literary works. By analyzing the image schema in classical poetry, we can understand the poet's perspective on issues and grasp the metaphorical expressions and the poet's thoughts and emotions in Tang poetry (Deng & Zhong, 2012:77). Below, we will briefly review the development of Image Schema Theory.

The concept of image schema originated from Lakoff & Johnson's (1980) Conceptual Metaphor Theory. The term "image" is often used in psychology as a mental representation, referring to the visual image of an object formed through imagination. German philosopher Kant believed that schema is a bridge between perception and concept, a means of establishing connections between concepts and objects, and a necessary process for creating image and meaning (Wang, 2006:172). Image schema has different definitions: 1) It is a dynamic structure that is constantly reproduced in perceptual interaction and sensory-motor activity, which provides coherence and structure to our experience (Johnson, 1987:xiv); 2) It can generally be defined as dynamic simulations of spatial relations and spatial motion (Gibbs & Colston, 1995:349). Li (2007) points out that there are many different expressions of the definition of image schema, but they are the same in essence. Firstly, as humans, we have our own bodies (or, cognitive embodiment), and we are always engaged in various activities: our bodies are always in contact and interaction with the external objective world. In these seemingly unrelated activities, image schema arises and makes these seemingly unrelated activities coherent, giving abstract activities a concrete structure, and laying the foundation for the further formation of "categories" and "concepts" in cognition. Therefore, we can use image schema to understand these activities, and connect seemingly unrelated activities.

### *2.4 Image Schema Theory and Poetry Translation*

Generally, two different language cultures often have different interpretations of the same image. This is also a common problem in poetry translation. Specifically, when readers of the target language try to understand the symbolic words and images in Chinese poetry, they often face three situations: first, the image schema of the source language and the target language coincide; second, the image schema of the source language and the target language differ; third, the image schema of the source language and the target language are missing (Deng & Zhong, 2012: 78). However, Chen (2010: 25-28) points out that "when the image schema is clearly shared, the grammatical structures of the two languages can be connected to each other in a shared conceptual system, and image transformation can proceed smoothly. When the image schema is not clearly shared, the grammatical structures of the two languages cannot find a shared image schema concept system, and then interpretive translation comes into play to make up for it." Therefore, the image schema theory can not only be used to guide poetry interpretation, but also to guide poetry translation, helping translators solve the difficulties of image schema translation.

This study aims to apply the image schema theory of cognitive linguistics to study the gains and losses of the English

translation of Mao Zedong's poetry, and this can help translators explore translation strategies and methods for translating Mao Zedong's poetry and other classical Chinese poetry.

### 3. Analyzing Translations of Dynamic Images in Mao Zedong's Poetry

There are different classification methods for image schema, but their core is based on human bodily experience. The basic image schema mainly includes: space, container, path, force, whole-part, center-periphery, etc. These classifications can also be combined to form more complex schema. In the following, we will focus on analyzing several cases of Mao Zedong's poetry that contain dynamic image schema, and study their corresponding translations. Starting from the theory of image schema, we will explore the gains and losses of these translations.

#### Translation Example 1.

Source Text:

红旗跃过汀江，  
直下龙岩上杭。  
收拾金瓯一片，  
分田分地真忙。（《清平乐·蒋桂战争》）

Source text Romanized:

Hóng qí yuè guò tīng jiāng ,  
Zhí xià àng yán shàng háng .  
Shōu shí jīn ōu yī piàn ,  
Fēn tián fēn dì zhēn máng . ( *Qīng Píng Yù è • Jiǎng Guì Zhàn Zhēng* )

This poem describes the war between reactionary warlords, which brings endless suffering to the people. The energy of the people's revolution is thus released, and the Red Army raises its banner and enters the Longyan-Shanghang area of Fujian Province to carry out land reform. The first line of this sentence uses the word "leap" to describe the movement of the red flag across the Ting River. On the surface, it describes the red flag, but in fact, it uses the red flag as a metonymy for the person holding the red flag, that is, the Red Army. The word "leap" originally means jumping, and with this word, it vividly depicts the fast marching speed of the Red Army, as if jumping, vividly reproducing the heroic fighting spirit of the Red Army soldiers leaping and charging into battle. Here are two versions of translation.

Target Text 1:

**Red banners leap over the Ting River  
Straight to Longyan and Shanghang.**

We have reclaimed part of the golden bowl,  
And land is being shared out with a will. (*Poems and Verses of Mao Zedong*, p.11)

Target Text 2:

**Over River Ting our red flags leap;  
To Longyan and Shanghang we sweep.**

A part of golden globe in hand,

We're busy sharing out the land. (*English Translation of Mao Zedong's Poems and Verses by Xu Yuanchong*, p.16)

This line describes a Red Army soldier holding a red flag and marching vigorously on a forced march, swiftly crossing the Ting River to establish a revolutionary base. Due to the rapid march, crossing the Ting River was like leaping over it. Wang (2006:182) points out that the English preposition "over" has four image schema, one of which expresses a trajectory passing a landmark to reach its destination. This is exactly the image schema that the poem wants to convey: the trajectory (in this case, the red flag) passes the landmark (the Ting River) and reaches its destination (Longyan and Shanghang on the other side of the river). Therefore, the preposition "over" combined with "leap" can vividly reproduce the dynamic image of "leaping" in the original text. However, from the perspective of whether the dynamic image schema chain is complete in the translation, Version 2 is slightly better than Version 1 for the following reasons.

In the first half of the poem, there is a line "Sowing on earth but grief and pain", which concretizes the abstract concept of "resentment" into a tangible thing that can be scattered and descend. In other words, there is a schematic framework that governs the entire poem: "the warlord melee leads to widespread resentment", including the related dynamic image of "scattering". This image is linked to other images of the same type in the poem, forming a chain of dynamic images. In the second half of the poem, after the Red Army captured Longyan and Shanghang, they eliminated feudal rule, liberated the people, and redistributed land, bringing about a new atmosphere. Therefore, there is another line "A part of golden globe in hand", which appears to be related to the previous "scattering" image. Therefore, "To Longyan and Shanghang we sweep" is not just a description of arriving at the destination, but also carries the meaning of liberating a place, eliminating the pain and resentment brought about by reactionary rule. Hence, the well-chosen verb "sweep" in Version 2 corresponds well with the image of "scattering". In the *Oxford Advanced Learner's Dictionary* (6<sup>th</sup> edition), "sweep" means "to remove something from a surface using a brush, your hand, etc." (p.1786), which coincides with the meaning of "cleaning up" in the original text. Therefore, Version 2 can better reproduce the chain of dynamic images in the original text.

### Translation Example 2.

Source Text:

山舞银蛇，原驰蜡象。（《清平乐·蒋桂战争》）

Source text Romanized:

Shān wǔ yín shé , Yuán chí là xiàng . ( *Qīng Píng Yuè • Jiǎng Guì Zhàn Zhēng* )

This couplet describes the scenery Mao Zedong saw after a snowfall while marching on the Qin-Jin plateau. The poet climbed to a high point and looked into the distance. As his gaze moved, the snow-covered mountains appeared to be dancing and charging like silver snakes and wax-colored elephants on the northern plains. Normally, the continuous hills and towering mountains are motionless, but with the movement of the poet's vision, they gain a sense of movement. Therefore, the verbs "dance" and "charge" are used in the poem to highlight the imagined movement. In other words, this couplet has a dynamic image schema in both lines.

Target Text 1:

The mountains dance like *silver snakes*,

And the highlands charge like *wax-hued elephants*. (*Poems and Verses of Mao Zedong*, p.45)

Target Text 2:

Mountains like *silver serpents* dancing,

Highlands like *waxy elephants* advancing. (*English Translation of Mao Zedong's Poems and Verses by Xu Yuanhong*, p.59)

The first half of this couplet describes this scene: the winding mountain ranges and the snow-covered mountains appear to be like a silver snake dancing on the ground. It reproduces the scenery that the poet saw in the distance and his rich imagination, which conforms to the dynamic image schema of the Source-Path-Destination. To produce a good translation, the translator should try to reproduce this dynamic image schema as much as possible in the translation. Both versions use the verb "dance" to reproduce the image of "mountains dancing like snakes", but whether the following adverb should be "like silver snakes" or "like silver serpents" is worth considering. On the one hand, although the word "serpent" in English refers to a "large snake," according to the *New Oxford English-Chinese Dictionary* (2<sup>nd</sup> edition), it also has several other meanings, such as "a biblical name for Satan" (p.2006) and "(figurative) a sly or treacherous person, especially one who exploits a position of trust in order to betray it" (p.2006). These cultural connotations may easily lead readers to unnecessary negative interpretations. Furthermore, the original phrase "silver snakes" mainly expresses the image of the mountains dancing like snakes, rather than referring to a specific type of snake. At the same time, according to the *Oxford English-Chinese Dictionary* (2<sup>nd</sup> edition) and *New Oxford English-Chinese Dictionary* (2<sup>nd</sup> edition), there is the phrase "snake dance" in English, which refers to a serpentine procession or dance performed to celebrate a victory or welcome alumni in Western schools. Therefore, Version 1 is better.

The second half of this couplet describes the scene after the snowfall: the peaks of the mountains on the Qin-Jin Plateau, like white elephants, charge and advance across the land. Similar to the first half, it gives a sense of movement to still objects. Therefore, in the translation, the verbs (or verb participle forms) "charge" and "advance" are used respectively. However, the word "charge" in the *Oxford Advanced Learner's Dictionary* (6<sup>th</sup> edition) is explained as "to rush forward and attack somebody or something" (p.266), which may be associated with violent connotations in cultural implications, leading English readers to have negative associations. Therefore, although both translations can reproduce the metaphor of movement in the original text, "advance" is more appropriate in cultural implications.

Although the original text uses the term "wax elephants", it is appropriate to translate it as "glittering elephants" or "shining elephants" in order to activate the image of "mountains like wax-hued elephants advancing" in the reader's mind. The poet's "wax elephants" describe the mountains after a heavy snowfall, shining brightly under the sun - these mountains, like elephants coated in wax, reflect the sunshine and sparkle. Therefore, the translator should emphasize the meaning of "like wax-coated elephants shining in the sun" in the original text, rather than translating the original text in a word-for-word way. At the same time, according to the *New Oxford English-Chinese Dictionary* (2<sup>nd</sup> edition), the word "wax" has other meanings, such as "a fit of anger" (p.2466), and the derived word "waxy" has meanings such as "angry, bad-tempered" (p.2466), so the word-for-word translation of the original term as "wax-hued elephants" or "waxy elephants" would not only be awkward but also could lead to misinterpretation by readers. Therefore, we can translate the couplet into "Mountains like silver snakes dancing, Highlands like glittering elephants advancing" based on the above analysis.

#### 4. Conclusion

As can be seen from the above analysis of translations, there is not just one way to translate the dynamic image in Mao Zedong's poetry and other Chinese literary works. To successfully translate the image in poetry and literature, it is essential to have an understanding of the shared image schema between the source and target languages. Translators can make use of the theory of image schema in cognitive linguistics and fully consider the allusions and cultural connotations behind the image in the original text during the translation process. Only by using these methods can Mao Zedong's poetry be translated effectively into the target language. As Mao Zedong's poetry is full of images, it is suggested that it be analyzed based on specific types and available translations. This study makes an attempt in this direction, and hope this can open up avenues for future studies.

#### References

- Chen, J. (2010). A Valuable Exploration of Cognitive Translation Research——A Review On Construal across Languages: A Cognitive Linguistic Approach to Translation. *Chinese Translators Journal*, 31(5), 25-28. [In Chinese: 陈吉荣. (2010). 认知翻译研究的可贵探索——评《跨越语言的识解: 翻译的认知语言学探索》. *中国翻译*, 31(5), 25-28.]
- Chi, M. (2011). Translation Strategies of Symbolism in Classical Chinese Poetry from the Perspective of Image Schema. *Journal of Mianyang Teachers' College*, 30(4), 75-77. [In Chinese: 迟梦筠. (2011). 从意象图式理论看中国古典诗词象征手法的翻译策略. *绵阳师范学院学报*, 30(4), 75-77.]
- Deng, J. & Zhong, S. (2012). Analysis on the Poem "On the Stork Tower" and Its Translations from the Perspective of Image Schema. *Journal of Jiaying University (Philosophy & Social Sciences)*, 30(9), 77-82. [In Chinese: 邓洁雯, 钟书能. (2012). 从意象图式视角解析《登鹤雀楼》及其翻译. *嘉应学院学报 (哲学社会科学)*, 30(9), 77-82.]
- Gibbs Jr, R. W. & Colston, H. L. (1995). The Cognitive Psychological Reality of Image Schemas and their Transformations. *Cognitive linguistics*, (6)4: 349
- Huang, S. (2019). English Translation of Chinese Classical Poetry from the Perspective of Cognitive Metonymy —— Taking Su Shi's Poetry as an Example. *Journal of Fujian Agriculture and Forestry University(Philosophy and Social Sciences)*, 22(1), 105-112. [In Chinese: 黄少静. (2019). 认知转喻视角下中国古典诗词英译研究——以苏轼诗词为例. *福建农林大学学报 (哲学社会科学版)*, 22(1), 105-112.]
- Johnson, M. (1987). *The body in the mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago press.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. University of Chicago press.
- Lakoff, G. & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor*. Chicago/London: The University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar, volume I: Theoretical prerequisites*. Stanford: Stanford University Press.
- Li, C., Li, J. & Li, W. (2008). On the English Translation of Allusion of Mao Zedong Poems. *Journal of Southwest Minzu University(Humanities and Social Sciences Edition)*, 29(S1), 1-4+2. [In Chinese: 李崇月, 李静, 李文. (2008). 毛泽东诗词中典故的英译. *西南民族大学学报(人文社会科学版)*, 29(S1), 1-4+2.]
- Li, F. (2007). On Image Schema Theory. *Foreign Languages and Literature*, 1, 1-2. [In Chinese: 李福印. (2007). 意象图式理论. *四川外语学院学报*, 1, 1-2.]
- Liu, G. (2014). A Research on the Construction of Image Schema in the Translation of Ancient Poems. *Journal of Jiaozuo University*, 28(1), 128-130. [In Chinese: 刘国立. (2014). 古诗词翻译中意象图式建构研究. *焦作大学学报*, 28(1), 128-130.]



- Long, J., Liu, M. & Liu, X. (2013). On Construction and Transmission of Culture Image in English Translation of Mao Zedong's Poems. *Journal of Hunan University of Science and Technology(Social Science Edition)*. 16(4), 17-20. [In Chinese, 龙君, 刘明东, 刘霞. (2013). 毛泽东诗词英译的文化意象的建构与传递. *湖南第一师范学院学报(社会科学版)*, 16(4), 17-20.]
- Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological review*, 99(4), 587-604.
- Qi, J. & Hu, H. (2020). Retrospection and Reflection on the Cross-Cultural Communication of Mao Zedong's Poems. *Journal of Xiangtan University(Philosophy and Social Sciences)*. 44(5), 6-12. [In Chinese, 齐静, 胡慧. (2020). 关于毛泽东诗词跨文化传播的回顾与思考. *湘潭大学学报(哲学社会科学版)*, 44(5), 6-12.]
- Turner, M. (1987). *Death is the mother of beauty: Mind, Metaphor, Criticism*. Chicago: The University of Chicago Press.
- Wang, P. (2007). Four English Version of the Word "Fly" in the Poems by Mao Ze-dong. *Foreign Language Research*. (05)121-123. [In Chinese, 王平. (2007). 传神达意 重在理解——评毛泽东诗词中含“飞”字句式的英译. *外语学刊*, (05)121-123]
- Wang, Y. (2006). *Cognitive Linguistics*. Shanghai Foreign Language Education Press. [In Chinese, 王寅. (2006). *认知语言学*. 上海外语教育出版社.]
- Yu, L. (2016). A Comparative Study on the Translation of Culture-loaded Words in English Versions of Mao Zedong's Poems. *Foreign Language Research*. (06), 106-109. [In Chinese, 余立霞. (2016). 毛泽东诗词英译本中文化负载词翻译的对比研究. *外语学刊*, (06), 106-109.]
- Zhang, H. & Jiang, C. (2021). On the Translator's Subjectivity from the Perspective of the Cognitive Linguistics—— Taking Niannujiao Theme: Reminiscing Olden Days at Chibi as an example. *English Square*, 10, 29-32. [In Chinese, 张涵, 姜成. (2021). 论认知语言学翻译观视角下译者主体性的体现——以《念奴娇 赤壁怀古》英译为为例. *英语广场*, 10, 29-32.]
- Zhu, X. (2005). The Sense of Reality Reflected by the Images in Poetry's Ideorealm of Song Dynasty: in Comparison with that of Tang Dynasty. *Journal of Fujian University of Technology*, (5), 473. [In Chinese: 朱晓慧. (2005). 宋词词境中的意象质感——兼与唐诗意象质感比较. *福建工程学院学报*, (5), 473. ]

A Brief Introduction of the Author: Liang Haowen (In Chinese, 梁昊文), born in Guangzhou, China in 1990, is now a lecturer of Guangzhou Xinhua University. Main Research Interests: Translation Teaching, Cognitive Translation Studies, Cognitive Linguistics.

# A ChatGPT-driven Study of Translating and Post-editing Chinese Enterprise Publicity Materials

Junchao WANG<sup>1</sup>

<sup>1</sup> *Guangdong University of Foreign Studies, Guangzhou, China 201510091@oamail.gdufs.edu.cn*

## Abstract

The China Import and Export Fair (a.k.a. Canton Fair), the first promotion platform for China's foreign trade, is a key topic for academic and business discussions. Driven by ChatGPT, the present study first constructs a bilingual database concerning the Canton Fair, and then compares the ChatGPT-generated and human-translated multimodal texts such as exhibition invitation letters and main visual images of the Fair. One of the findings is that artificial intelligence can help translate in a top-down way, that is, from a macroscopic and three-dimensional knowledge framework to language usage at a microscopic level, which is different from the previous bottom-up translation practice, teaching, and research. In addition, the comparison of different translations can also provide empirical rules and insights for writing, translating, and post-editing enterprise publicity materials in the new artificial intelligence era.

**Keywords:** artificial intelligence; Canton Fair; ChatGPT; Chinese enterprise publicity; post-editing; AI translation

## 1. Introduction

In 2023, ChatGPT, the milestone of AI, has become the hottest new trend in the entire digital ecosystem. It is an intelligent chatbot developed by the American artificial intelligence laboratory OpenAI, which is powered by Large Language Models (LLM), new human-machine interaction interfaces, key large-scale algorithms, big data, and large computing power (Zhu & Wang, 2023: 113-115). It is powerful in terms of language generation, in-context learning, and world knowledge storage. Within less than a week of its launch, it has attracted over one million users and attention from such fields as language education, academic research, and medical surgery. This study takes the post-pandemic publicity of the China Import and Export Fair (Canton Fair) as an example and makes attempts to explore new paths and methods for writing, translating, and post-editing Chinese enterprise publicity in the new era of technology, to shed light on the teaching, learning and practicing of business translation and editing. To do so, I will first build a bilingual database concerning the Canton Fair based on the chats with ChatGPT to provide knowledge for the translation, post-editing, and understanding of the source texts, then analyze the advantages and disadvantages of ChatGPT-generated translations by contrasting them with those rendered by professional translators, and put forwards suggestions on improving the translation and editing of enterprise publicity materials.

## 2. Reviewing ChatGPT and Translation Studies

### 2.1 State-of-the-art of ChatGPT

Different from traditional professional-generated content (PGC) and user-generated content (UGC), AI-generated content (AIGC) is based on human-machine interaction and created by artificial intelligence to produce the content one needs, with features such as intelligent production, user customization, and "cross-border" ecology (Zhao & Zhao, 2023: 2). Among them, "cross-border" ecology refers to the fact that AI-generated content has broken through the barrier of language and text as a single modality, and can create content with a wide range of genres and modalities, including codes, texts, images, audios, videos, and games. For example, ChatGPT can not only create various literary genres such as novels, poetry, speeches, advertising copy, papers, and programming, but also has mature development in fields such as painting creation, medical diagnosis, and strategic management.

The debate over whether the new generation of artificial intelligence is "empowering" or "dangerous" has sparked discussions (*The Lancet Digital Health*, 2023; Yadava, 2023; Wu, 2023). Elon Musk, one of the co-founders of OpenAI (left in 2018), expressed concern about its development, tweeting, "ChatGPT is scary good. We are not far from dangerously strong AI." Artificial General Intelligence (AGI), also known as Strong or human-level AI, can learn in any environment or scenario and act as an intelligent robot or system capable of rational action (Goertzel, 2014: 2-3). Existing research is focused on two main aspects: infrastructure and application. The former focuses on the ontology of artificial intelligence, such as whether AI is an evolution or a revolution (Gordijn & Henk, 2023) and analysis of the three basic elements - data, algorithms, and computing power (Min, 2023; Xu & Yan, 2023). The latter focuses on content production and scenario layout, with domestic research focusing on language education, publishing, and information security (e.g., Wu & Huang, 2023; Zhang, et al., 2023), while international studies highlight medical diagnosis (Chiwon, 2023) and scientific research (e.g., Dowling & Brian, 2023). We found only two related articles on "ChatGPT" and "translation" on the China National Knowledge Infrastructure (CNKI) (i.e., Wang & Liu, 2023; Cui & Li, 2023), leaving ChatGPT-based translation studies unexplored.

### 2.2 Reviewing Chinese Enterprises' Overseas Publicity

With the growth of China's national strength, overseas publicity of the People's Republic of China and its enterprises has become more important, as it has an impact on China's national image and corporate image, and exerts cultural influence on the international community. In the article *Telling the Story of China Well through Foreign Language Translation*, Yin (2020) pointed out that "in the new era, language research should not be confined within the language itself, but actively connect with national strategy, integrate into national strategy, and focus on the fundamental issues at the national strategic level." As part of telling the story of China well, Chinese enterprises' overseas publicity has significant significance for enterprises going global and enhancing China's international influence. Therefore, in recent years, research on the translation and image construction of Chinese enterprises has become a hot topic, with results showing diverse theoretical methods, domestic and international contexts highlighting, local and regional foci, strategic translation issues, diverse principles and methods, and clustered research objects. However, systematic research based on the corpus or big data is rarely found, with the exception of Wang's discussions (2019; 2022a; 2022b).

### **3. Building a Bilingual Database of the Canton Fair based on Chatbot**

#### *3.1 A Web of Knowledge on the Fair*

By chatting with ChatGPT, we can collect the background knowledge required for translation practice or research. In addition, the Chatbot has powerful translation capabilities and can translate the information provided by customers or collected by ChatGPT. With the help of machine-assisted translation software, related information can be aligned bilingually to provide background knowledge reserves for translation. For example, all MT apps translated "中国对贸易中心" as "China Foreign Trade Center", but ChatGPT is very powerful and translates it as "China Foreign Trade Centre," which is the official English name on its website. Before translating the publicity materials of the Fair, we first have a simple Q&A with ChatGPT (as shown in Figure 1) to collect relevant background information and Chinese-English expressions related to the enterprise. The background information is collected based on the AI-given answers to the following 15 questions.

- (1) Please tell me more about "中国对外贸易中心".
- (2) Please write an introduction to Canton Fair in China.
- (3) Please introduce Canton Fair.
- (4) Would you say something more about the Fair?
- (5) Please talk more about the exhibits and products at the Canton Fair.
- (6) Please tell me the number of participants and exhibitors in the Fair over 66 years.
- (7) Please let me know the contributions of the Canton Fair, locally and globally.
- (8) Could you tell me the history of the Fair?
- (9) Please tell me about the design and development of the four exhibition complexes of the Canton Fair over the past 66 years.
- (10) Do you know the slogan of the Canton Fair over the years?
- (11) Please tell me the logo design of the Canton Fair over the past 66 years.
- (12) Please tell me the varieties of products exhibited at the Canton Fair over 66 years.
- (13) Please tell me the investments and incomes of the Canton Fair over 66 years.
- (14) Please collect the announcement of each Canton Fair over the past 66 years.
- (15) How Canton Fair went on during the COVID-19 pandemic?

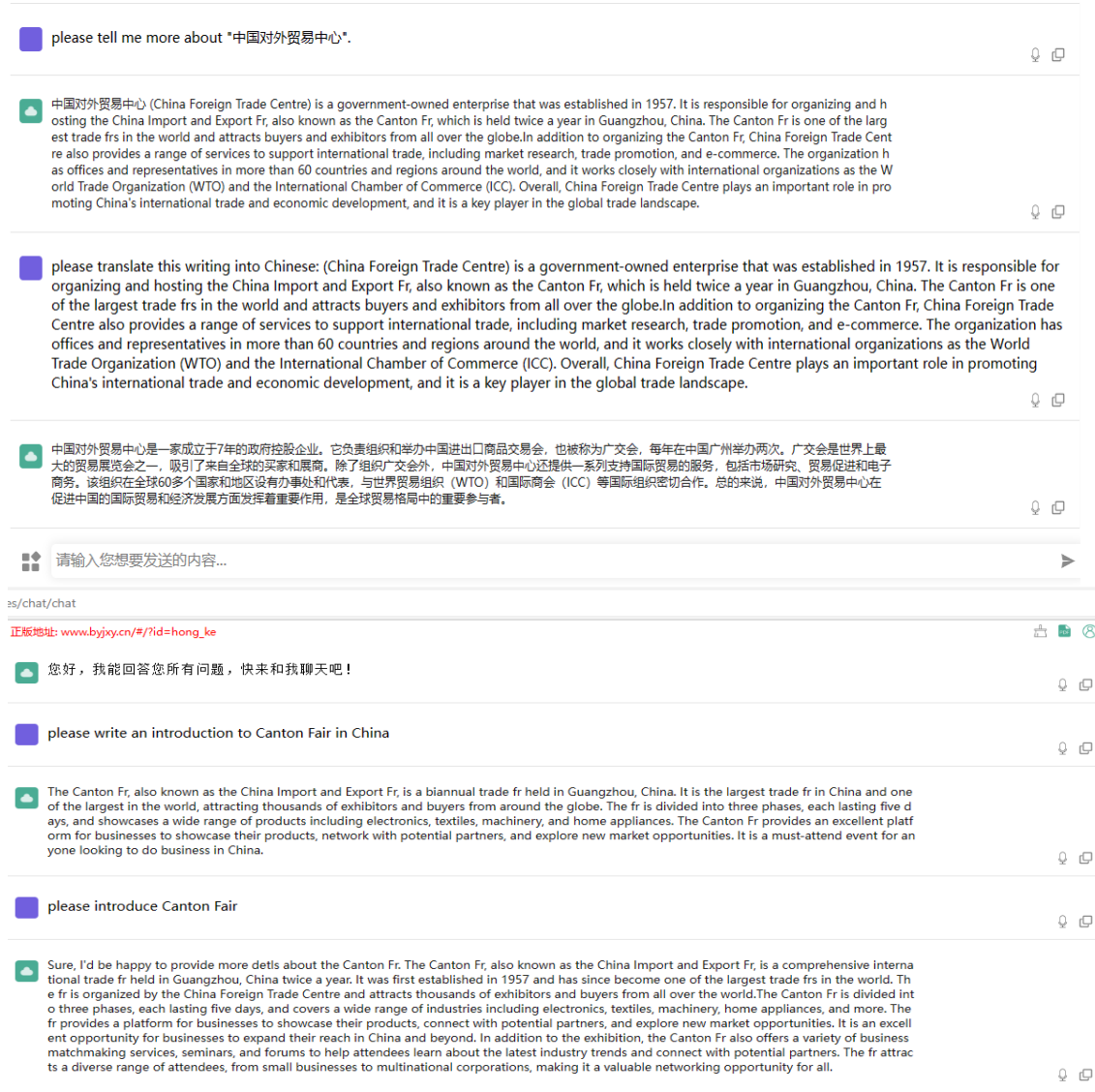


Figure 1. Human-machine Chat

Based on the data, we can construct a background knowledge web or framework (frame) related to the translation/research object, namely, "Canton Fair", covering the enterprise background, its development history, conference arrangements, exhibitor booths, scales, and numbers of participants, functional contributions, exhibition hall constructions, exhibition slogans, logo design, types of exhibits, its inputs and revenues, exhibition opening, and organizational models during the epidemic (see Figure 2).

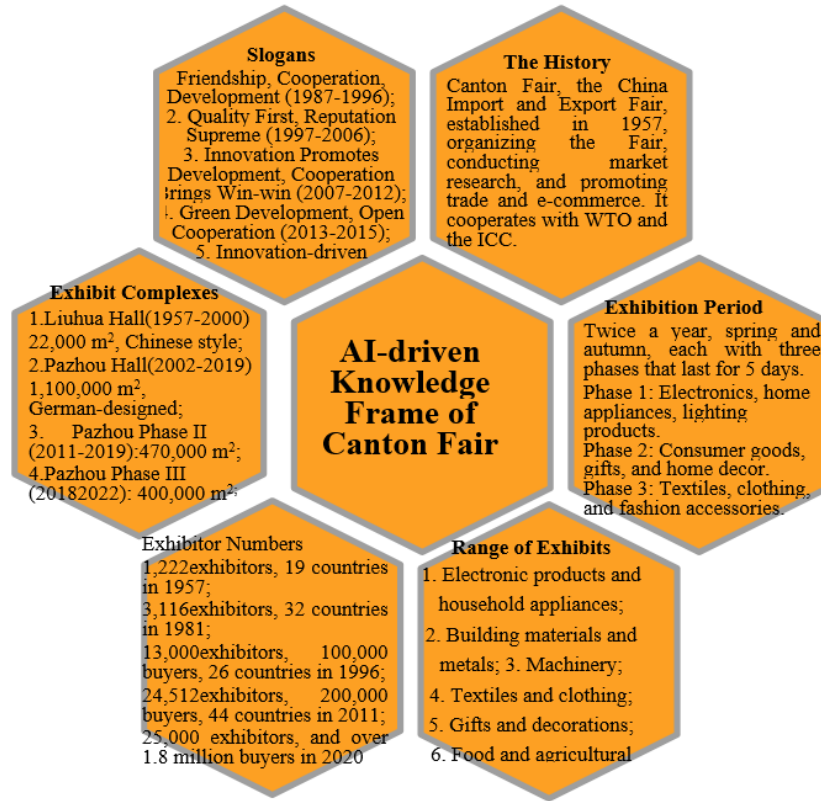


Figure 2. Constructing a Knowledge Frame of Canton Fair Driven by AI

### 3.2 Terminology Database Based on AI-generated Texts

Under the background of big data and artificial intelligence, terminology translation has been closely integrated with information technology, and terminology management has emerged (Tao & Zhang, 2016: 83). Terminology management is the processing of terminology expressions, knowledge, and resources involved in a project to serve a certain practical activity, usually including "collecting, describing, processing, storing, editing, presenting, searching, maintaining, and sharing terminology" (Wang, 2013: 11). Common terminology systems at home and abroad include SDL MultiTerm, Acrolinx IQ Terminology Manager, MultiQA, TBX Maker, STAR TermStar, TermFactory, Snowman, and other terminology tools (Wang & Wang, 2019: 11). There are generally three sources for creating a terminology database: the client provides a terminology list, the translator builds a terminology base through network resources, and the translator organizes terminology based on the original files (Kang, et al., 2015: 108). The terminology management in this article is not based on terminology conversion, terminology annotation, or terminology extraction from the source texts, but uses ChatGPT to generate texts related to the Canton Fair (5000+ words), with the "Text Graph Version 3.0" program (<https://www.wenziyun.cn/>) to generate high-frequency terminology and perform word frequency analyses.

With a word frequency analysis, we further remove function words and punctuations, obtaining the high-frequency words and their frequencies, including Fair (140 times), Canton (99 times), products (35 times), exhibitors (34 times), trade (30 times), buyers (23 times), features (15 times), session (14 times), range (12 times), businesses (11 times), held (11 times), complex (11 times), phase (10 times), machinery (9 times), textiles (9 times), appliances (8 times), electronics (8 times), (7 times), Pazhou (7 times), attendees (6 times), participants (6 times), ceramics (5 times), offline (5 times), and in-person (4 times). Based on this, we can present the Canton Fair terminology cloud in a 3D visualization (see Figure 3).



I see I listen I enjoy //

Tóngqī jǔbàn: guóji zhìhuì xiǎnshì jí xitǒng jíjǐ zhǎn (Shēnzhèn)

Guóji biāozhì jí LED zhǎ (Shēnzhèn)

Zhǔbàn dānwèi: Guǎngzhōu jiāoyì huì guǎnggào yǒuxiàn gōngsī /

Zhōngguó duìwài mào y guǎngzhōu zhǎnzhǎi yǒuxiàn gōngsī /

Shēnzhèn shì hǎi zhǎn huì zhǎn kējì yǒuxi zérèn gōngsī

Xiébàn dānwèi: Zhōngguó guāngxué guāngdiǎnzǐ hángyè xiéhuì fāguāng èr jídiǎn xiǎnshì yǐngyòng fēnhuì / Zhōngguó gōngyè hézuò xiéhuì shēng guāng shì xùn zhuānyè wěiyuánhui

Target text (ChatGPT): Audio-Visual Fusion Technology and Application Exhibition (Shenzhen) And China Audio-Visual Industry Technology and Application ConferenceIoT Ecology Leads the FutureShenzhen World International Convention and Exhibition Center (Baoan New Hall) I see I listen I enjoy Concurrently held: International Smart Display and System Integration Exhibition (Shenzhen) International Signage and LED Exhibition (Shenzhen) Organizers: Guangzhou Trade Fr Advertising Co., Ltd. / China Foreign Trade Guangzhou Exhibition Corporation / Shenzhen Haizhan Exhibition Technology Co., Ltd. Co-organizers: China Optics and Optoelectronics Industry Association LED Display Application Branch / China Industrial Cooperation Association Audio-Visual Professional Committee (“/” added by the author)

Target text (Human): **ISLE 2022**

**Lighting Acousto Video Technology and Application Exhibition (Shenzhen) & Lighting Acousto Video Industrial Technology and Application Conference**

#### **IOT Ecology AI Technology**

Shenzhen International Convention and Exhibition Center (Bao'an New Hall)

#### **I see I listen I enjoy**

**Co-exhibitions:** International Smart Display & Integrated System Exhibition (Shenzhen)

International Sign and LED Exhibition (Shenzhen)

**Organizers:** Guangzhou Trade Fair Advertising Co., Ltd.

China Foreign Trade Guangzhou Exhibition Co., Ltd.

Shenzhen Haizhan Exhibition Technology Co., Ltd.

**Co-organizers:** Association of Luminescent Diode Display Application of COEMA, Lighting Acousto Video Committee of CIC (translated by the author)

As for information grouping, this source text consists of three parts: the exhibition name, the exhibition theme/slogan, and the exhibition organizer. The first translation (by ChatGPT) seems to be accurate. However, there are a few translation issues: Firstly, as there is no punctuation at the end of the original text, ChatGPT cannot recognize the meaning groups, resulting in a non-running translation. Secondly, the exhibition theme or slogan in the original text is symmetrical with four characters and a rhyme. Yet, ChatGPT translated it as "IoT Ecology Leads the Future", losing the original catchy feature. Thirdly, some of the exhibition organizers' names are proper nouns, and some already have English expressions. However, ChatGPT's translation did not follow the principle of "being named after its owner" for proper noun translation. For example, the official website name of the China Lighting Acousto Video Industry is "CLAV", which is an abbreviation. In addition, the length of the translations should also be considered for names. For example, the translation "China Industrial Cooperation Association Audio-Visual Professional Committee" is too long. Fourthly, the original text is written in both Chinese and English, and ChatGPT mechanically gives redundant expressions, such as translating "Shenzhen World 深圳国际会展中心" as "Shenzhen World International Convention and Exhibition Center".

Taking these translation issues into account, we suggest some improvements for ST writing and post-editing. (1) Attention should be paid to the use of punctuation and to the clear division of sentence groups in terms of ST writing and polishing. (2) Creative expressions in the source language are the focus that translators should pay attention to in the machine translation post-editing (MTPE) process. (3) Proper nouns can be expressed with the official English expressions or the ones already in use in ST, or be checked and verified during the post-editing. (4) The original text should be written without redundant expressions in both Chinese and English to avoid redundant or even erroneous translations.

#### *4.2 AI and Human Translation of the Main Visual Image*

The main visual image refers to the first thing people see and the image is designed based on that visual impression to meet the visual needs of the target audience and the designer. The main visual image of the Canton Fair is a commercial logo/design that provides a lot of information and intended meanings. For example, the 127<sup>th</sup> Canton Fair was held online due to the pandemic, so the “Golden Bond”, with the slogan “广交云上互利天下 Canton Fair, Global Share”, was designed as an



exclusive visual image for the event. Similarly, that for the 130<sup>th</sup> Canton Fair is composed of the English name of the Fair, "130", the logo, and the period (1957-2021), highlighting the theme of promoting domestic and international mutual connections and the first-time integration of online and offline for the Fair. Here, we take the visual image of this year's visual image and text as an example.

The main visual image is a multimodal text that is information-based and descriptive, mainly describing the designer's concept and creative information of multimodal wholes (including text, numbers, graphics, colors, etc.). Below, we compare and evaluate the AI and translator's translations of the source text (see below), focusing on the cross-culture and cross-language transfer of creative information and how AI/ChatGPT processes such information.

Source text: 倒计时 50 天! 第 133 届广交会主视觉形象精彩亮相!  
盼望着, 盼望着, 春天来了, 广交会的脚步近了。  
值此 50 天倒计时之际, 第 133 届广交会主视觉形象正式发布!



Figure 4. The main visual image of the 133<sup>rd</sup> Canton Fair

Source: <https://www.cantonfair.org.cn/>

——以见礼遇, 大尊显故也《后汉书·礼仪志上》

第 133 届广交会主视觉形象由广交会品牌标识、主题口号、广交会展馆形象、红色礼盒造型组成。

礼盒造型位于视觉中心, 寓意第 133 届广交会作为我国疫情防控优化转段以来首个重启线下的超大规模国际经贸盛会, 是为推动全球经济贸易复苏和发展献上的一份开放合作的“大礼”。

礼盒顶部飘逸舒展、流畅自然的金边缎带巧妙地勾勒出数字“133”和广交会展馆 ABCD 四个区域的标志性造型, 寓意第 133 届广交会首次启用 D 区新展馆, 迈上四馆同步举办的新征程, 生动展现了“中国第一展”的崭新形象。

礼盒扣结是一枚熠熠生辉的金色宝相花, 代表广交会的功能和作用。广交会创办 66 年来, 始终坚守初心使命, 为服务国际贸易、促进内外联通、推动经济发展作出了重要贡献。由此延伸的红色丝带, 寓意着越来越多的中国企业从广交会启航, 昂首迈向国际市场, 全球客商在这里与中国结缘, 分享机遇, 共同发展。

画面主色调选取红和金, 寓意广交会传承红色基因、再创金色荣耀。在各方支持下, 第 133 届广交会将实现办展规模新突破、功能拓展新突破、品质提升新突破, 为加快建设贸易强国、推动共建“一带一路”高质量发展、维护多元稳定的国际经济格局和经贸关系、构建人类命运共同体作出新的更大贡献。

Source text Romanized: dǎojìshí wǔshí tiān! dì yī sān sān jiè guǎngjiāohuì zhǔshì jù é x íngxi àng jīngcǎi liàngxiàng!  
pànwàng zhe, pànwàng zhe, chūntiān lái le, guǎngjiāohuì de jiǎobù jìn le. zhí cǐ wǔshí tiān dǎojìshí zhī jì, dì yī sān sān  
jiè guǎngjiāohuì zhǔ shì jù é x íngxi àng zhèngshì fābù!

—— yǐ jiàn lǐ yù, dà zūn xiǎn gù yě, hòuhànshū · lǐyí zhì shàng

dì yī sān sān jiè guǎngjiāohuì zhǔ shì jù é x íngxi àng yóu guǎngjiāohuì pǐnpái biāozhì, zhùtí kǒuhào, guǎngjiāohuì  
zhǎnguǎn xiàngxiàng, hóngsè lǐhé zàoxíng zǔchéng.

lǐhé zàoxíng wèi yú shì jù é zhōngxīn, yùyì dì yī sān sān jiè guǎngjiāohuì zuò wéi wǒ guó yì qíng fāng kòng yōu huà  
zhuàn duàn yǐ lǐ shǒu gè chóng qǐ xiànxia de chāo dà guīmò guóji jīngmào shèng huì, shì wèi tuīdòng quánqiú jīngjì  
màoyì fāzhǎn hé fāzhǎn xiàng shàng de yīfèn kāifāng hézuò de "dàlì".

lǐhé dǐngbù piāoyī shì yī yìyìshēnghuī de jīnsè bǎoxiàng huā, dàibiāo guǎngjiāohuì de gōngnéng hé zuò yì àng. guǎngjiāohuì  
chuàngbàn liùliù nián lái, shǐzhōng jiānshǒu chūxīn shǐmìng, wèi fúwù guóji màoyì, cùjìn nèiwài liántōng,



tuīdòng jīngjì fāzhǎn zuòchū le zhòngyào gòngxiàn. yóucǐ yánshēn de hóngsè sīdài, yùyì zhe yuè yuè duō de zhōngguó qǐyè cóng guǎngjìōuhuì qǐ háng, ángshǒu màixiàng guójiè shìchǎng, quánqiú kèshāng zài zhèlǐ yǔ zhōngguó jiéyuán, fēnxiǎng jīhuì, gòngtóng fāzhǎn.

huàmiàn zhǔsèdiào xuǎnqǔ hóng hé jīn, yùyì guǎngjìōuhuì chuánchéng hóngsè jīyīn, zài gèfāng zhīchí xià, dì yīsān sān jiè guǎngjìōuhuì jiāng shíxiàn bànzǎn guīmò xīn túpò, gōngnéng tuòzhǎn xīn túpò, pǐnzhì tíshēng xīn túpò, wèi jiākuài jiànshè màoyì qiángguó, tuīdòng gòn jiàn "yī dài yī lù" gāo zhìliàng fāzhǎn, wéihù duōuán wèndì guójiè jīngjì gǒu jí hé jīngmào guānxi, gòujiàn rénlèi míngyòng tǐ zuòchū xīn de gèng dà gòngxiàn.

Target text (ChatGPT): Countdown 50 days!

The main visual image of the 133<sup>rd</sup> Canton Fair is stunningly released! Looking forward, looking forward, spring is coming. The Canton Fair is getting closer. On the occasion of the 50-day countdown, The main visual image of the 133<sup>rd</sup> Canton Fair is officially released!

"After Han Shu, Volume of Rites and Ceremonies"

The main visual image of the 133<sup>rd</sup> Canton Fair consists of the Canton Fair brand logo, theme slogan, Canton Fair exhibition hall image, and red gift box shape. The gift box shape is located at the visual center, symbolizing that the 133<sup>rd</sup> Canton Fair is the first large-scale international economic and trade event to restart offline since China's epidemic prevention and control optimization, and is a "big gift" for promoting global economic and trade recovery and development through open cooperation.

The flowing and natural golden satin ribbon cleverly outlines the iconic shapes of the numbers "133" and the ABCD four areas of the Canton Fair exhibition hall on the top of the gift box, symbolizing that the 133<sup>rd</sup> Canton Fair has started a new journey with the simultaneous operation of the four halls, vividly demonstrating the new image of "China's No. 1 Exhibition".

The gift box buckle is a glittering golden lotus flower, representing the function and role of the Canton Fair. In the 66 years since its establishment, the Canton Fair has always adhered to its original mission and made important contributions to serving international trade, promoting internal and external connectivity, and promoting economic development. The extending red ribbon symbolizes that more and more Chinese companies are setting off from the Canton Fair, heading towards the international market with their heads held high, and global merchants are connecting with China here, sharing opportunities, and developing together.

The main color of the picture is red and gold, symbolizing that the Canton Fair inherits the red genes and creates a new golden glory. With the support of all parties, the 133<sup>rd</sup> Canton Fair will achieve new breakthroughs in exhibition scale, function expansion, and quality improvement, and make new and greater contributions to accelerating the construction of a strong trading nation, promoting the high-quality development of jointly building the "Belt and Road", maintaining a diverse and stable international economic pattern and economic and trade relations, and building a community with a shared future for mankind. (underlined by the author)

Target text (Human): **COUNT DOWN: 50 days!**

### **The main visual image of the 133rd Canton Fair**

With great anticipation, spring is coming. With greater anticipation, Canton Fair is around the corner. On this day, the main visual image of the 133<sup>rd</sup> Canton Fair is officially released.

"It is a sign of great honor by showing respect to whomever you meet", reads "On Rites (I)" of *The History of Later Han*.

This visual image consists of the logo of Canton Fair, the slogan, the image of the exhibition complex, and a bright gift box.

The box, sitting at the center, carries the message that the 133<sup>rd</sup> Canton Fair, as the first super large-scale event in terms of international economy and trade to be restarted onsite since Covid-19, is indeed a gift for co-promoting the recovery and development of global economy and trade.

The golden-rimmed ribbon, naturally folded on the top, is designed in the iconic shape of 1-3-3 and the four halls of the complex, implying that Exhibition Hall D is to be used for the first time during this year's Fair. This marks that Canton Fair embarks on a new journey of holding grand exhibitions and vividly renews its image as "China's First Exhibition".

The buckling knot made in the pattern of a shining flower called **Boxiang Flower** or Fortune & Fairy Flower represents the functions and roles of Canton Fair. Since its establishment in 1957, Canton Fair has always adhered to its original missions and made great contributions to serving international trade, promoting international connectivity, and

economic development. And the red ribbon is designed to be extending from the flower, symbolizing that more Chinese enterprises are setting sail from Canton Fair and heading towards the international market and that global customers tie the bond with China, share opportunities, and seek co-development at Canton Fair.

The major colors of the image are red and **golden**, which are employed to convey that Canton Fair inherits the “red gene” (i.e., the tradition of revolution) and creates a new golden glory. With all the support, the 133rd Canton Fair will achieve breakthroughs in terms of the exhibition scales, expanding functions, and quality improvement, and make greater contributions to accelerating the construction of trading power, promoting high-quality development under the Belt and Road Initiative, maintaining a multi-dimensional and stable international economic pattern and trade relations, and building a Community of Shared Future for Mankind. (translated by the author)

With Grammarly, both translations received a score of 86, which is much lower than the score of the translation of the exhibition invitation letter. This indicates that creative (multimodal) texts are more difficult for both artificial intelligence and translators. Specifically, there were several translation issues in the translation: (1) Repetition in the original Chinese text resulted in repetitive translations by the ChatGPT, such as "looking forward, looking forward" and the translation of 亮相 and 发布; (2) The translation of long and complex noun phrases in Chinese (as the subject) resulted in a list of nouns in the translation, such as 礼盒造型 translated as "The gift box shape", 礼盒顶部飘逸舒展、流畅自然的金边缎带 translated as "The flowing and natural golden satin ribbon", and 礼盒扣结 translated as "The gift box buckle"; (3) There were literal translations of individual characters, such as 金 in 主色调选取红和金 being translated as "gold", which refers to the material, while "golden" should be used to refer to the color; (4) There were issues with the translation of culture-loaded words or expressions, such as the Canton Fr's logo -宝相花- being directly translated as "lotus flower", losing its cultural connotations. As one of China's traditional patterns, 宝相花 is also known as 宝莲花 and, along with 摇钱树 and 聚宝盆, is referred to as the "Three Treasures of Good Fortune", representing the meanings of treasure, immortality, and good fortune (as shown in Figure 5). Its English name is commonly known as "Composite Flowers", but we believe that culture-loaded words are one of the channels for exporting Chinese culture. Therefore, we adopted a "transliteration + translation" method based on its cultural connotations, namely, "Boxiang Flower or Fortune & Fairy Flower".



Figure 5. Pictures of Boxiang Flower

Source: <https://image.baidu.com/>

Although the professor translator spent a lot of time dealing with translation difficulties and post-editing, the translated text is similar in length and number of words and sentences to that by ChatGPT, but it is more readable (37:29, readability). Based on the above analysis, three suggestions can be made for AI translation and post-editing. (1) Pre-translation processing. For shorter texts, Chinese expressions such as noun phrases, undetermined meanings of individual characters, repetitive expressions, and running sentences should be processed before using artificial intelligence translation. (2) Post-editing, especially for longer texts where pre-translation processing may not be feasible, can improve quality and efficiency by revising and editing the four main issues mentioned above. (3) There should be a focus on improving the translation of culture-loaded words, with an emphasis on the effectiveness of the dissemination of Chinese cultural elements and ideas.

## 5. Conclusion

With the end of the pandemic, China's enterprises make an effort to go global by expanding internal and international publicity. Rather than use ChatGPT to translate corporate publicity materials, the present study attempts to illustrate how new AI or Chatbot can be used for building a bilingual knowledge pool and a terminology database for a certain translation project and make a comparing and contrastive analysis of the AI-generated and human translated texts. It finds that artificial intelligence can help translate in a top-down way, that is, from a macroscopic and three-dimensional knowledge framework to language usage at a microscopic level, which is different from the previous bottom-up translation practice, teaching, and research.

Besides, the comparison of AI- and human-produced translations can provide empirical rules and insights for writing, translating, and post-editing the (enterprise publicity) materials in the new artificial intelligence era.

#### Declarations and Acknowledgment:

The author declares there is no conflict of interest.

This work was supported by the National Social Science Fund Key Project (under grant number 20&ZD312), the Ministry of Education's Humanities and Social Sciences Fund (under grant number 22YJCZH250); and the Key Project of Guangdong University Journal Research Association (under grant number 20220201).

#### References

- Ahn, C. (2023). Exploring ChatGPT for information of cardiopulmonary resuscitation. *Resuscitation*, 185, 109729.
- Cui, L., & Li, S. (2023). The subjectivity of translator in knowledge translation studies. *Foreign Language Electronic Teaching*, 2: 78-81+94+116. [In Chinese: 崔凌霄, 李淑华. 译者主体性的知识翻译学阐释. 外语电化教学]
- Dowling, M., & Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama conjecture. *Finance Research Letters*, 53, 103662.
- Goertzel, B. (2014). Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 5(1): 1-46.
- Gordijn, B., & Henk, T. H. (2023). ChatGPT: Evolution or revolution? *Medicine, Health Care, and Philosophy*, 26(1): 1-2.
- Kang, N., Wei, Y., & Qu, L. (2015). Creation, application, and management of terminology database in a computer-aided translation environment. *Journal of Qingdao University of Science and Technology (Social Sciences)*, 31(3): 107-110. [In Chinese: 康宁, 尉莹, 曲乐. 计算机辅助翻译环境下术语库创建、应用与术语管理. 青岛科技大学学报(社会科学版)]
- Min, H. (2023). The dual impact of ChatGPT on computing power. *Communication World*, 8: 4. [In Chinese: 闵海兰. ChatGPT 对算力存在双重影响. 通信世界]
- Tao, L., & Zhang, B. (2016). A microscopic study on Chinese terminology translation research: An interview with professor Zhang Boran. *Foreign Language Research*, 33(2): 83-86. [In Chinese: 陶李春, 张柏然. 中国术语翻译研究探微——张柏然教授访谈录. 外语研究]
- The Lancet Digital Health. (2023). ChatGPT: Friend or foe? *The Lancet Digital Health*, 5(3): E102.
- Wang, H. (2013). On terminology management in practice. *Chinese Science and Technology Terminology*, 2: 11-14. [In Chinese: 王华树. 浅议实践中的术语管理. 中国科技术语]
- Wang, H., Liu, S. (2023). Research on intelligent translation education: Concept, path, and trend. *Shanghai Translation*, 3:47-51+95. [In Chinese: 王华树, 刘世界. 智慧翻译教育研究: 理念、路径与趋势. 上海翻译]
- Wang, H., Wang, S. (2019). Terminology management in translation scenarios: Process, tools, and trends. *Chinese Science and Technology Terminology*, 21(3):9-14. [In Chinese: 王华树, 王少爽. 翻译场景下的术语管理: 流程、工具与趋势. 中国科技术语]
- Wang, J. (2019). Construction of a research framework for Chinese enterprise's going global publicity translation. *Shanghai Translation*, 2: 62-66. (In Chinese: 王俊超. 构建中国企业“走出去”外宣翻译的研究框架——基于 500 强企业网页外宣语料库. 上海翻译)
- Wang, J. (2022a). A contrastive study of multimodal advertising in China-foreign automobile enterprises and the translation strategies. *Journal of Asia-Pacific and European Business*, 2: 14-20.
- Wang, J. (2022b). The translation and promotion of Chinese national drinks brands from the perspective of advertisement history. *Communication across Borders: Translation & Interpreting*, 3: 1-8.
- Wu, G. (2023). Artificial general intelligence: Empowerment or danger. *People's Forum*, 5: 48-52. [In Chinese: 吴冠军. 通用人工智能: 是“赋能”还是“危险”. 人民论坛]
- Wu, W., & Huang, H. (2023). Intelligent creation, deep integration and ethic crisis: A new exploration of the application of ChatGPT in the digital publishing industry. *Chinese Editor*, June: 1-5. [In Chinese: 吴炜华, 黄珩. 智能创作、深度融入与伦理危机——ChatGPT 在数字出版行业的应用前景新探. 中国编辑]
- Xu, J., & Yan, R. (2023). On the evidence attribute of large language model materials: Taking and Wenxin Yiyuan as examples. *Journal of Sichuan Normal University (Social Sciences)*:1-15. [In Chinese: 徐继敏, 严若冰. 论大语言模型材料的证据属性——以 ChatGPT 和文心一言为例. 四川师范大学学报(社会科学版)]

- Yadava, Om Prakash. (2023). ChatGPT: A foe or an ally? *Indian Journal of Thoracic and Cardiovascular Surgery*, 39(3): 217-21.
- Yin, B. (2020). Doing well in foreign propaganda translation and telling China's story Well[OL]. China Daily Website, <https://cn.chinadaily.com.cn/a/202006/17/WS5ee9db8ea31027ab2a8d0a9b.html> [In Chinese: 尹铂淳. 做好外宣翻译, 讲好中国故事. 中国日报网]
- Zhang, C., Weng, F., & Zhang, Y. (2023). Application, status, and trend of ChatGPT in the field of network security. *Journal of Information Security Research*, 9(6): 500-509. [In Chinese: 张弛, 翁方宸, 张玉清. ChatGPT 在网络安全领域的应用、现状与趋势. 信息安全研究]
- Zhao, Z., & Zhao, Y. (2023). ChatGPT: The era of intelligent media for language and writing. *Southern Media Studies*, 1: 1-2. [In Chinese: 赵子忠, 赵一然. ChatGPT : 语言文字的智媒体时代. 南方传媒研究]
- Zhu, G., & Wang, X. (2023). Operating Mode, Key and Future Vision of ChatGPT. *Journal of Xinjiang Normal University (Philosophy and Social Sciences Edition)*, 44(4):113-122. [In Chinese: 朱光辉, 王喜文. ChatGPT 的运行模式、关键技术及未来图景. 新疆师范大学学报(哲学社会科学版)]

**Author Bio:** Junchao WANG is currently an associate editor of the Journal of Guangdong University of Foreign Studies, with a Ph.D. in Translation Studies. Her research interest lies in translation history, business translation, and metaphor, with a publication of over 30 articles on translation studies and foreign language teaching in peer-reviewed journals.

# A Historical Comparative Study on the Annual Reports of Chinese Public Companies

Chaowang REN<sup>1</sup>; Haolin ZHENG<sup>2</sup>

<sup>1</sup> *Guangdong University of Technology, Guangzhou, China paulren@gdut.edu.cn*

<sup>2</sup> *Guangdong University of Technology, Guangzhou, China 2112316004@mail2.gdut.edu.cn*

## Abstract

The significant changes in human social history will undoubtedly lead to alterations in language use. Concurrently, the diachronic study of language change can reflect historical changes in society. This paper employs the principles of corpus construction and utilizes AntConc and other corpus retrieval tools to conduct a diachronic comparative analysis of discourse variation in management discussion and analysis. This analysis is based on a self-built corpus and integrates findings from diachronic linguistics and sociolinguistics. We examine discourse variation in management discussions through three key levels: word frequency, keywords, and collocates. By delving into the text, we perform a diachronic comparative analysis of management discussions, emphasizing the dialectical relationship between discourse and society. This approach highlights language use across different time periods, reveals patterns of language change, and supports the development and transformation of society. The authors assert that corpus-based diachronic language change elucidates the nature of language evolution as a reflection of social change. Furthermore, it provides valuable insights for a deeper understanding of the complex causes of language variation and offers material for the standardization of language expression.

Keywords: a corpus-based approach, historical comparative study, language variation, annual reports

## 1. Introduction

Diachronic language research, a key aspect of usage-based linguistics, emphasizes the importance of utilizing real corpora. The rapid development of diachronic studies can largely be attributed to the ongoing construction of electronic diachronic corpora. In linguistic literature, this area of research is commonly referred to as "historical linguistics" or "diachronic linguistics," focusing on the mechanisms of language evolution (Campbell, 1999: xiv). Relevant studies encompass evidence of lexical evolution (etymology), the construction of language family trees (contrastive linguistics), and the exploration of mechanisms that facilitate the transformation of content words into functional words (grammaticalization).

Furthermore, there is a notable overlap between diachronic language studies and the field of sociolinguistics. Diachronic variation within sociolinguistics, archaic usage in dialectology, and contemporary language interpretation through the lens of language contact all exhibit diachronic characteristics. Language is inherently dynamic, and change is fundamental to its nature. Thus, conducting diachronic language research is essential within the field of linguistics.

The term "Management's Discussion & Analysis" (MD&A) denotes the management's evaluation and interpretation of a company's past operational performance, as well as forward-looking assessments concerning future trends, as presented in periodic reports by publicly listed companies. Often regarded as the "core and soul" of financial reporting, MD&A constitutes an essential element of the information disclosed by these entities. It provides insights that extend beyond the GAAP-compliant data found in tables and notes, thereby fulfilling the growing demand among investors for pertinent and prospective information.

This disclosure mechanism allows investors to better understand management's perceptions of significant historical events, assess key risks, and grasp the rationale behind projections regarding future developments. Due to its essential function in improving the quality of corporate information transparency, MD&A has been widely adopted in nations such as the United States, the United Kingdom, and Canada. Notably, it was recognized as one of the top ten financial reporting challenges by the American Association of Financial Managers in 2005. Additionally, in response to the Enron accounting scandal, the U.S. Securities and Exchange Commission instituted MD&A as a crucial corrective action. In China, the China Securities Regulatory Commission launched the MD&A framework in 2002 and subsequently reinforced regulations in 2005.

A comprehensive analysis of the discourse within annual reports of Chinese public companies can significantly enhance investors' understanding of these entities, thereby facilitating more effective resource allocation in the capital market. Moreover, a comparative study of MD&A sections in the annual reports of Chinese public companies, grounded in corpus analysis, has the potential to enrich diachronic research in China. This approach will expand the corpus field and allow for a multidimensional and multi-level examination of the linguistic features present in these reports. Consequently, this research

aims to systematically and objectively explore the characteristics of language change within the annual reports of public companies in China, as well as elucidate the dialectical relationship between language change and societal factors.

## 2. Literature review

Corporate annual reports represent a critical component of the information disclosed by publicly listed companies, playing an essential role in enhancing the quality of corporate information disclosure. In linguistic literature, diachronic language research is commonly referred to as "historical linguistics" or "diachronic linguistics," focusing on the mechanisms underlying language variation.

To refine and mature our Management's Discussion & Analysis (MD&A) system, there is an increasing necessity to improve the quality of information disclosure among listed companies. The academic community must urgently engage in theoretical research on MD&A to provide foundational support and guidance for future practices in this area.

### 2.1 Previous studies on the discourse of Management's Discussion and Analysis

Management's Discussion and Analysis (MD&A) serves as an essential element of a public company's annual report or quarterly filing, focusing on the organization's performance. In this section, the management team and executives—commonly referred to as the C-suite—offer an analysis of the company's performance employing both qualitative and quantitative metrics. Furthermore, the MD&A may include discussions related to compliance, risk factors, and future strategies, encompassing organizational objectives and upcoming initiatives.

MD&A constitutes the central component of the "Business Situation Discussion and Analysis" section in the annual reports of publicly listed companies. It incorporates essential financial and non-financial information, along with both historical and prospective data, thereby significantly enhancing the disclosure of corporate financial information.

Numerous studies have advanced the understanding of the effects and implications of MD&A. Caserio (2019) explored whether financial companies in the United States exhibited a tendency to manipulate the tone of their MD&A to engage in impression management practices. Lee (2019) investigated how the financial expertise of audit committees affected the quality of the textual information presented in the MD&A sections of corporate annual reports. Durnev (2020) analyzed the robustness of disclosures within the MD&A, concluding that firms in industries characterized by lower entry costs, larger scales, and less substitutable products were more likely to provide comprehensive disclosures. Jaspersen (2020) utilized supervised machine learning models to predict earnings management behaviors among property and casualty insurers based on their MD&A submissions. Lastly, Pajuste (2020) contended that increased scrutiny from market participants in more liquid companies motivated managers to obscure unfavorable financial outcomes through heightened complexity in reporting.

### 2.2 Previous corpus-based diachronic discourse analysis

Conrad (2002) offered an extensive examination of methodologies within corpus linguistics that pertain to discourse-level phenomena. The primary aim was to amalgamate two analytical perspectives: performing a thorough discourse analysis of individual texts while also generalizing insights across the entire corpus. Biber (1998) distinguished two principal approaches to this endeavor: 'top-down' and 'bottom-up'. The study employed relatively large corpora characterized by parallel structures and content from various contemporary periods, specifically utilizing the SiBol corpora to monitor changes in modern language use in relation to associated social, cultural, and political transformations as evidenced in linguistic expression.

Partington (2010) provided a comprehensive overview of corpus-assisted discourse studies (CADS) and multidimensional CADS (MD-CADS), with the objective of clarifying the methodologies typically employed in this area and offering insights into newspaper corpora. Brown et al. (2012) analyzed a corpus of Medieval Spanish texts to explore the factors affecting the outcome of the initial /f/ in Latin FV- words as they evolved into Modern Standard Spanish. Li (2013) introduced a discourse-level joint model for argument identification and role assignment, thereby advancing the understanding of inter-sentence arguments within discourse.

The challenges of achieving stable and reliable results from corpus analysis, particularly in studies of lexical variation, were highlighted by Miller (2015), who illustrated these difficulties through experiments analyzing a corpus related to university-level introductory psychology textbooks. Odebrecht (2017) introduced a multi-layer corpus architecture featuring multiple tokenizations, utilizing the open-source historical, diachronic corpus of German known as Register in Diachronic German Science. Furthermore, speeches delivered from 1950 onwards have often been made available online through video recordings, enabling new research avenues in multimodal linguistics. This includes studies on the integration of speech and gesture in meaning construction, as well as analyses of recurrent gestures and multimodal constructions (Debras, 2021).

### 2.3 Research questions

(1) What linguistic variations can be identified in the Management's Discussion and Analysis section of annual reports from Chinese public companies across different time periods?

(2) What are the key factors contributing to the linguistic variations observed in the Management's Discussion and Analysis sections of annual reports from Chinese public companies?

### 3. Data and methodology

By collecting and cleaning the target texts, we obtained a refined version to construct the Chinese corpus. Analytical tools were utilized to extract the linguistic features from the corpus.

#### 3.1 Data

##### 3.1.1 Data collecting

In this study, we excluded ST companies and those with incomplete data from the 145 Chinese firms listed in the 2022 Fortune Global 500. Ultimately, we selected the management discussion and analysis (MD&A) sections of annual reports from 30 A-share companies for the years 2007 to 2021 as our research data. All data were sourced from [www.cninfo.com.cn](http://www.cninfo.com.cn). According to annual report disclosure guidelines, companies must include specific MD&A information in the "Report of the Board of Directors," such as: an overview of operational performance during the reporting period; a list of main business activities by product and region; an analysis of significant changes in operating results and profit composition compared to the previous year; statements on the impact of major changes in the production environment, macro-policies, and regulations on financial position and operating results; the board's business plan for the upcoming year; and the profit forecast for the New Year. Although some companies label their MD&A sections as "business situation discussion and analysis," the content is fundamentally similar. Therefore, these sections were also included in the study's data set.

##### 3.1.2 Data cleaning

Data cleaning is to keep useful data in the corpus and delete noise data. First, the annual reports of 30 target companies were obtained in pdf from [www.cninfo.com.cn](http://www.cninfo.com.cn). Then the pdf file was converted into Word version, and the management discussion and analysis section of the annual report was manually selected to delete the table content, English characters and other information that affected the search results.

#### 3.2 Methodology

This chapter introduces the research methodology and analytic tools employed in this research. First a brief introduction to corpus and corpus-based approach is presented. Then the issues of corpus construction are discussed, including the principles, data selection and annotation. Finally, it is followed by the description of the analytic tools.

##### 3.2.1 An introduction to corpus and corpus-based approach

Corpus is from the Latin word, meaning "body", hence a corpus refers to any body of text. But the term "corpus" has the specific connotation in modern linguistics.

According to Sinclair (1999), corpus is defined as "a collection of naturally occurring language text, chosen to characterize a state or variety of a language." Biber et al. (1998) emphasize the authenticity and purpose of a corpus by noting "It is a large and principled collection of natural texts." It is used to describe a collection of naturally occurring examples of language, consisting of written texts or tape recordings; recently it has been reserved for collections of texts that are stored and accessed electronically (Hunston, 2002).

There are two main approaches to corpus linguistics, a corpus-based approach and a corpus-driven approach. In the corpus-based approach, corpora are used mainly to "expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study". In contrast, corpus-driven approach claims that 'the theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus (Tognini-Bonelli, 2001:65-84). The most notable difference is that the corpus-driven approach claims to be a new paradigm within which a whole language can be described, while the corpus-based approach has made no such claim. The corpus-based approach is widely applied in nearly all branches of linguistics. Therefore, the corpus-based approach has made it possible to conduct new kinds of investigations into language use.

The advantages of corpus-based approach are obvious. The corpus-based approach takes advantage of computers' capability for fast, accurate, and complex analysis; the extensive information about language use found in large collections of natural texts from multiple registers; and the rich descriptions that result from integrating quantitative findings and functional interpretations (Biber et al., 1998:233). Even though the corpus does not contain new information about language, the corpus-based approach offers a new perspective view with the help of software.

##### 3.2.2 Constructing the MD&A corpus

At the "2003 International Conference on Corpus Linguistics" held in Shanghai Jiaotong University, John Sinclair mentioned that corpus construction became a popular practice among language teachers and researches. Although there are some

established-corpora available, yet those do not always suit the purpose of every research. Researchers can establish a mini-corpus with the availability of computer technology, software and on-line data to satisfy the aim of research.

As the purpose of this paper is to discuss the diachronic linguistic variation of the Management's Discussion and Analysis in annual reports of Chinese public companies during the 15 years, the Management's Discussion and Analysis in annual reports is selected as data for the construction of MD&A Corpus.

Allan Bell (1991) suggests that when establishing a news corpus, "Decisions are needed on three questions: the time period the sample will cover, the days to be sampled within that period, and the specific issues or bulletins to be sampled within those days."

For selecting data from MD&A in annual, random sampling is employed. The selected annual reports are from 2007 to 2021. As most of the websites can be restored as a text file, the easiest way of obtaining texts is downloading the electronic format from the internet.

### 3.3 Analytical procedures

The AntConc was used to retrieve and sort high-frequency words in the three subcorpora. Functional words refer to the words that only play a grammatical role without inflection and have no practical meaning in the literature, such as prepositions, articles, conjunctions, pronouns, auxiliary verbs, etc. They cannot reflect or represent the content of the literature, and their resolution is very low or even zero. Therefore, the analysis of this paper will exclude the function words and only analyze the content words. In this paper, content words with the top ten word frequency in the three subcorpora will be selected for analysis. If necessary, the collocate search function of AntConc will be combined for joint analysis and the search results will be sorted out in Excel, so as to obtain the similarities and differences of words in the three subcorpora. Two analytic tools were used.

#### (1) AntConc

AntConc, developed by Professor Laurence Anthony at Waseda University, is a free corpus retrieval tool widely used in corpus linguistics, translation studies, and foreign language teaching (Anthony, 2024). It is recognized as an authoritative software for linguistic research and supports various languages, including Chinese, Japanese, and Korean. AntConc can perform corpus statistics on UTF-8 and UTF-16 encoded texts and utilizes regular expressions, particularly beneficial for processing Chinese corpora, where it can batch-add spaces.

#### (2) Microsoft Excel

Microsoft Excel, part of the Microsoft Office suite, is a spreadsheet software utilized for data processing, statistical analysis, and decision-making. It is extensively used in management, finance, and statistical fields. In this paper, Excel will be employed for data statistics and organization.

## 4. The Diachronic Linguistic Variations of the Corporate Annual Reports in China

Corpus-based approaches to language study have introduced new dimensions to linguistic description and to various applications by permitting some degree of automatic analysis of texts. In this study, the self-established MD&A Corpus is used to describe discourse linguistic variation from wordlist level, keywords level, and collocation and word cluster level. In the following sections, MD&A Corpus is researched with the aid of the AntConc tool from both quantitative and qualitative perspectives.

### 4.1 The General Description of the MD&A Corpus

As mentioned in Chapter Three, the corpus for the current study is compiled based on the principles to guarantee the balance thus the validity of analysis result, it is still necessary to present some of the general information to verify it. MD&A Corpus is composed of three sub-corpora, with 26312315 words and 39755 types in total. The six sub-corpora line up chronologically. The general statistics of each sub-corpus are listed in Table 4-1.

Table 4-1 General statistics of three sub-corpora

Text file	File size	Tokens (running words)	Types (distinct words)	Type/ token ratio (TTR)	Standardized TTR	Mean word length
2007-2011	6724990	806660	16972	2.26	37.45	1.9
2012-2016	8794900	1044206	20661	2.13	38.55	1.92
2017-2021	10792425	1268889	26830	2.26	40.41	1.92



Type Token Ratio (TTR): is a measure of the ratio of different words to the total number of words in a text, sometimes used as a measure of the difficulty of a 32 passage or text. It is also referred to as lexical density which is normally expressed as a percentage and is calculated by the formula:

$$\text{Lexical Density} = \frac{\text{The number of different words}}{\text{The total number of words in the text}} \times 100$$

Standardized Type Token Ratio (STTR): is an average type/token ratio based on consecutive 1,000-word chunks of text, which can alleviate the influence of the number of tokens in the sub-corpus.

From Table 4-1, we can get the general view of the three sub-corpora. Even though the construction of corpus is based on the rules, yet there are some differences between each sub-corpus because of the different formats of each year. The tokens of the 3 sub-corpora are 806660, 1044206, and 1268889 respectively. It can also be found that the TTR varies from 2.13 to 2.26. Different numbers of tokens of each year influence the figure of TTR, thus, so STTR is more meaningful to show that the lexical diversity is comparatively stable during the past 15 years. The STTR of the 3 sub-corpora are 37.45, 38.55, and 40.41 respectively. The language difficulty of MD&A maintains the same during the 15 years.

#### 4.2 The Diachronic Description of the Linguistic Features of the Subcorpora

The MD&A Corpus consists of three sub-corpora. When these sub-corpora are combined and analyzed, they reveal a new perspective. The total size of the MD&A Corpus is 26,312,315 words, with a total of 39,755 unique types. This chapter will explore the linguistic variation within the MD&A Corpus over the past 15 years by analyzing the three sub-corpora across different dimensions.

To conduct a word frequency analysis, we utilized the Wordlist function in AntConc. This tool counts all words in the corpus and presents them in an ordered list, allowing us to identify the most frequent terms. The frequency of these words reflects linguistic variations over the 15-year period. Once the Wordlist is generated, we can assess the frequencies. Additionally, the MD&A Corpus exhibits patterns in parts of speech that align with findings from other corpora. This section will provide statistical analyses and comparative insights into the top ten high-frequency nouns, verbs, and adjectives derived from the three sub-corpora.

##### 4.2.1 The high-frequency nouns in the subcorpora

A noun is a word that represents a person, a place, or a thing. A thing in this definition can be a physical entity or it can be an abstract idea. A noun may be used as a subject of a sentence, as a direct object, as an indirect object, or it may be used as the object of a preposition (Quirk & Greenbaum, 1973:49). Thus, the nouns need to be discussed firstly. The top 10 nouns in each sub-corpus are listed in Table 4-2.

Table 4-2 The top 10 nouns in three sub-corpora

Rank	High-frequency nouns in 2007-2011	Frequency	High-frequency nouns in 2012-2016	Frequency	High-frequency nouns in 2017-2021	Frequency
1	公司 (Company)	13056	公司 (Company)	18293	公司 (Company)	16826
2	业务 (Business)	9842	业务 (Business)	13684	风险 (Risk)	14304
3	管理 (Management)	7850	管理 (Management)	11650	管理 (Management)	13322
4	风险 (Risk)	6925	风险 (Risk)	11078	业务 (Business)	13246
5	市场 (Market)	5380	资产 (Assets)	7881	金融 (Finance)	8615
6	人民币 (Renminbi)	5229	金融 (Finance)	6440	客户 (Client)	8227
7	资产 (Assets)	4428	市场 (Market)	6279	服务 (Service)	8113
8	银行 (Bank)	4180	银行 (Bank)	6205	资产 (Assets)	7647
9	集团 (Group)	4005	客户 (Client)	6127	集团 (Group)	7274
10	产品 (Product)	3921	集团 (Group)	5921	产品 (Product)	6985

From the words in Table 4-2, we can see that the focus of MD&A is about company and the annual reports serves as a window for the whole world to see the company. Through comparison, it is found that the top 10 nouns in the word frequency ranking of the three sub-corpora do not change very much. The word frequency ranking of most words still remains in the top 10 of the sub-corpora under the influence of time, but there are also some notable phenomena. For example, the word "公司 (company)" always ranks in the top one words frequencies of the three sub-corpora, which indicates that the part of

management's discussion and analysis focuses on the company and always focuses on discussing and analyzing all affairs related to the company.

Secondly, in the comparison, it is found that among the top ten words in the corpus in terms of word frequency in each time period, the word "管理(management)" is also ranking the third space in each sub-corpora. From this phenomenon, it can be concluded that the management of the company pays more attention to the management of the company and promotes the good operation of the company by improving the company's management system, so as to better attract investors to invest in the company.

In addition, we can see that the word frequency of the word "风险(risk)" ranked 4th in the two sub-corpora from 2007 to 2011 and 2012 to 2016, but jumped to the second place in the 2017-2021 sub-corpus. We can speculate that more and more enterprises are paying more attention to the standardization of enterprises and the perfection and improvement of investment system. In particular, since 2019, the impact of the novel coronavirus pneumonia epidemic has increased the downward risk of the global economy, which will also increase the business risk.

Moreover, it is worth noting that the word "服务(service)" suddenly appear in the third sub-corpus. In particular, the noun frequency of the word "服务(service)" ranks seventh in the third sub-corpus. After comparison, it is found that the word frequency ranking of "服务(service)" increases year by year in the three sub-corpora, which indicates that enterprises pay more attention to the improvement of service and pay attention to the provision and improvement of service. The frequency of "服务(service)" in each sub-corpus is listed in Table 4-3.(Note: This data is the ranking and frequency of "服务(service)" in the three sub-corpora, regardless of speech.)

Table 4-3 The frequency of "服务(service)" in each sub-corpus

Period	Frequency	Rank
2007-2011	2,919	24
2012-2016	5,520	18
2017-2021	9,444	8

#### 4.2.2 The high-frequency verbs in the subcorpora

Table 4-4 The top 10 verbs in the three sub-corpora

Rank	High-frequency verbs in 2007-2011	Frequency	High-frequency verbs in 2012-2016	Frequency	High-frequency verbs in 2017-2021	Frequency
1	增长 (Grow)	4514	增长 (Grow)	5733	增长 (Grow)	5863
2	增加 (Increase)	3849	增加 (Increase)	4140	提升 (Enhance)	5659
3	实现 (Achieve)	2518	实现 (Achieve)	3757	实现 (Realize)	4762
4	提高 (Enhance)	2017	提升 (Enhance)	3714	增加 (Increase)	3599
5	提升 (Elevate)	1902	推进 (Advance)	2614	推进 (Promote)	3538
6	加强 (Strengthen)	1854	加强 (Strengthen)	2541	加强 (Strengthen)	2764
7	推进 (Advance)	1597	发展 (Develop)	2043	提供 (Provide)	2545
8	发展 (Develop)	1582	优化 (Optimize)	1926	优化 (Optimize)	2497
9	继续 (Continue)	1439	下降 (Decline)	1889	发展 (Develop)	2174
10	下降 (Decline)	1386	提高 (Improve)	1875	推动 (Drive)	2103

As can be seen from Table 4-4, the top ten verbs with word frequency in the three sub-corpora have not changed much except their ranking. However, we can also find that the top ten verbs with the highest frequency in the three sub-corpora have a common feature, that is, most of them express positive attitudes. Pava and Epstein (1993) thought most companies, in the disclosure of forward-looking information, tended to disclose good news not only for the most part, and the number of accuracy would more than double that of bad news. Many companies clearly have poor operating conditions, but if you only look at the MD&A, you will think that the situation of the company is promising.

#### 4.2.3 The high-frequency adjectives in the subcorpora

Table 4-5 The top 10 adjectives in three sub-corpora

Rank	High-frequency adjectives in 2007-2011	Frequency	High-frequency adjectives in 2012-2016	Frequency	High-frequency adjectives in 2017-2021	Frequency
1	主要 (Major)	1772	新 (New)	2646	新 (New)	3950
2	新 (New)	1432	有限 (Limited)	2192	综合 (Comprehensive)	1887
3	有限 (Limited)	825	主要 (Major)	2013	大 (Large)	1817
4	相关 (Relevant)	718	综合 (Comprehensive)	1440	主要 (Major)	1544
5	综合 (Comprehensive)	657	大 (Large)	1171	高 (High)	1379
6	宏观 (Macro)	630	相关 (Relevant)	1062	相关 (Relevant)	1309
7	总 (Total)	607	净 (Net)	839	有限 (Limited)	1278
8	大 (Large)	536	重大 (Significant)	748	绿色 (Green)	1107
9	不良 (Adverse)	495	总 (Total)	711	重大 (Significant)	964
10	净 (Net)	494	不良 (Adverse)	686	总 (Total)	899

As can be seen from Table 4-5, the top 10 adjectives in the three sub-corpora have not changed much except their ranking. However, it is worth noting that the word "green" popped up in the third sub-corpus and ranked seventh.

Since the beginning of the 21st century, the report to the 17th National Congress of the Communist Party of China has listed ecological civilization as a new requirement for the goal of building a well-off society in an all-round way. The report to the 18th CPC National Congress made important breakthroughs in the overall layout of the cause of socialism with Chinese characteristics, including ecological civilization in the five-in-one overall layout of the cause of socialism with Chinese characteristics. At the Fifth Plenary Session of the 18th CPC Central Committee in 2015, the concept of innovative, coordinated, green, open and shared development must be firmly established. Taking the improvement of ecological environment as the starting point, this paper systematically expounds the policies, measures and measures to promote our green development. The report of the 20th National Congress of the Communist Party of China put forward: "To promote green development and promote the harmonious coexistence between man and nature", "we must firmly comb and practice the concept that lucid waters and lush mountains are gold and silver mountains, and plan development from the height of the harmonious coexistence between man and nature. As the country gradually improves and attaches more importance to the concept of green development, listed companies also conform to the trend of The Times and the national development strategy to promote green development and achieve high-quality development of the economy and society.

#### 4.3 Discussions on the Reasons of Diachronic Linguistics Variations of the Corporate Annual Reports

The variations in linguistic characteristics of the Management's Discussion and Analysis (MD&A) sections over the past 25 years have been described and summarized with the aid of AntConc. Given the complexity of language change, this paper provides a preliminary explanation for these variations.

In examining high-frequency words, it is important to note that the Chinese language has a limited number of synonyms for certain noun expressions. This inherent characteristic restricts the precision and specificity of terms used in annual reports, which serve as formal documents reflecting a company's financial performance and related information. Consequently, companies must present their content with clarity and accuracy. In contrast, Chinese verbs often possess multiple expressions conveying similar meanings. To avoid redundancy, managers frequently employ varied verbs to articulate the same ideas. Furthermore, the influence of contemporary social conditions necessitates that corporate high-frequency terms evolve in tandem with societal changes and national policies.

Overall, the transformation of discourse in the MD&A sections of corporate annual reports can be attributed primarily to textual features, the companies' disclosure tendencies, and broader social changes

## 5. Conclusion

This study investigates the discourse linguistic variation in the Management's Discussion and Analysis (MD&A) sections of corporate annual reports from 2007 to 2021, focusing on five-year intervals. Through word frequency analysis, significant linguistic changes over the 15-year period have been identified.

The relationship between language and society is bidirectional; language both reflects and influences societal developments (Wardhaugh, 1986: 23). Dittmar (1976: 238) emphasizes that speech behavior and social structure are in constant interaction, with language serving as a mirror to societal evolution.

To comprehensively explore these linguistic changes, the research utilizes the MD&A Corpus, which consists of the MD&A sections from 30 corporate annual reports compiled at five-year intervals. The analysis employs AntConc software to establish a framework for diachronic examination of language shifts over the past 15 years. The study initiates with a literature review on diachronic studies and corpus-based discourse analysis, both domestically and internationally. Chapter 3 outlines the research methodology and the corpus development, while Chapter 4 provides a detailed description of the established corpora and summarizes key findings across three analytical levels. From the word frequency analysis, notable terms such as 公司 (enterprise), 风险 (risk), 管理 (management), and 业务 (business) indicate fluctuations in frequency that correspond to evolving corporate priorities over time. This shift reflects the changing emphasis of companies in response to broader societal dynamics. The rapid advancement of the internet in China has enabled stakeholders to access more direct and immediate corporate information, further contextualizing these findings.

This study is pioneering in its examination of discourse variation in Chinese corporate annual reports, offering fresh perspectives for future research. By collecting and comparing linguistic data over a 15-year span, the study visualizes historical trends through tables, creating a comprehensive overview akin to geological interpretations of earth's evolution through rock strata. Given that much of the existing research on historical corpora focuses on English and adheres to frameworks established by Western scholars, this work contributes significantly to the field of modern and contemporary Chinese by providing a diachronic comparative analysis of corporate discourse.

Ultimately, this research highlights the intricate nature of language change and its reflection of societal transformations. While some aspects of these changes remain difficult to explain, the study lays foundational insights for further exploration into the complex causes of language evolution and standardizes expressions within corporate communication. Future inquiries should continue to examine the multifaceted phenomena of language change, enhancing our understanding of its role in reflecting societal shifts.

#### **Declarations and Acknowledgment:**

The author declares there is no conflict of interest.

This work was supported by the Guangdong Provincial Fund for the Philosophy and Social Science (under grant number GD21CWY05).

#### **References**

- Allan, B. (1991). *The Language of News Media*. Oxford: Basil Blackwell Ltd.
- Anthony, L. (2024). *AntConc (Version 4.3.1)* [Computer software]. Waseda University. Available from <https://www.laurenceanthony.net/software>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Brown, E. L., & Raymond, W. D. (2012). How discourse context shapes the lexicon: Explaining the distribution of Spanish F-/H- words. *Diachronica*, 29(3), 297-317.
- Campbell, L. (1999). *Historical Linguistics: An Introduction*. Cambridge, MA: MIT Press.
- Caserio, C., Panaro, D., & Trucco, S. (2019). Management discussion and analysis: A tone analysis on US financial listed companies. *Management Decision*, 57(6), 1287-1304.
- Conrad, S. (2002). Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics*, 22, 143-158.
- Debras, C. (2021). How to prepare the video component of the diachronic corpus of political speeches for multimodal analysis. *Research in Corpus Linguistics*, 9(2), 129-145.
- Dittmar, N. (1976). *Sociolinguistics: A Critical Survey of Theory and Application*. London: Edward Arnold.
- Durnev, A., & Mangen, C. (2020). The spillover effects of MD&A disclosures for real investment: The role of industry competition. *Journal of Accounting and Economics*, 69(1), 101344.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Jaspersen, J. G., Richter, A., & Zoller, S. (2020). Predicting earnings management from qualitative disclosures. *S&P Global Market Intelligence Research Paper Series*. Retrieved from <https://www.spglobal.com>

- Lee, J., & Park, J. (2019). The impact of audit committee financial expertise on management discussion and analysis (MD&A) tone. *European Accounting Review*, 28(3), 475-498.
- Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 8-29). London: Longman.
- Li, P., Zhu, Q., & Zhou, G. (2013). Joint modeling of argument identification and role determination in Chinese event extraction with discourse-level information. *IJCAI*. Retrieved from <https://ijcai.org>
- Miller, D. G., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, 20(3), 275-298.
- Odebrecht, C., Belz, M., Zeldes, A., Lüdeling, A., & Krause, T. (2017). RIDGES herbology: Designing a diachronic multi-layer corpus. *Language Resources and Evaluation*, 51(3), 755-779.
- Pajuste, A., Poriete, E., & Novickis, R. (2020). Management reporting complexity and earnings management: Evidence from the Baltic markets. *Baltic Journal of Management*, 15(1), 85-104.
- Pava, M. L., & Epstein, M. J. (1993). How good is MD&A as an investment tool? *Journal of Accountancy*, 175(3), 51-53.
- Quirk, R., & Greenbaum, S. (1973). *A University Grammar of English*. London: Longman.
- Sinclair, J. (1999). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Co.
- Wardhaugh, R. (1986). *An Introduction to Sociolinguistics*. New York: Basil Blackwell.
- Author Bio:** Chaowang REN is currently an associate professor of Guangdong University of Technology, with a Ph.D. in Applied Linguistics. His research interest includes corpus linguistics, discourse analysis, and metaphor, with a publication of over 20 articles on corpus-based discourse studies in peer-reviewed journals. Haolin ZHENG is currently a postgraduate of Guangzhou University of Technology, majoring in translation and interpreting. His research interest lies in corpus-based translation and machine-aided translation studies.



## The Editorial Team of MLLL

The Editorial Committee, comprised of eight members, meets twice a year to validate the editorial process and take related decisions (topics, sections, guest editors, choice of reviewers). The head of this committee is elected by its members for a two-year renewable mandate. His or her job is to monitor the Journal's editorial process and website at regular intervals, and to pursue contacts with contributors and authors, as well as with the publishers and the distributor.

The Editorial Committee ensures and implements the process of assessing scholarly articles according to the double-blind principle. It also appoints referees for each issue from within the Reviewing Committee, and takes decisions on special contributions or those concerning specific sections (reviews, testimonies, tributes, etc.).

### **Editor-in-chief:**

- Xinliang JIANG, PhD in linguistics (Newcastle University), Toronto University, Canada

### **Editorial team:**

- Dongpin HU, PhD in Educational Technology (Expected 2023, The University of Hong Kong), Master's in General Educational Studies (The Education University of Hong Kong), Hong Kong
- Tony Whitt , Adjunct Faculty, College of Arts and Sciences, Roosevelt University; Instructor and Tutor at Various Colleges; Podcast Host and Producer; Master of Fine Arts (MFA) from Louisiana State University; Bachelor of Arts (B.A.), English Language and Literature from Western Michigan University
- Simin TANG, Linguistics PhD candidate at Saint Paul University Philippines; Associate Professor, Deputy Dean of the School of Foreign Languages, Jiaying College
- Rebecca Yunusa WALKER, PhD in Linguistics and English Language (Newcastle University UK), Masters in English Language and Linguistics (Newcastle University UK), UK
- Zhaofeng JIANG, Professor (Nanning Normal University), MA (University of Worcester, UK), translator and interpreter for China -ASEAN Expo, expert in intercultural communication.
- Anthony Bayer, Master's Degree from the University of Auckland along with a Secondary Teachers College Diploma from Auckland Secondary Teachers College, now part of the University of Auckland, New Zealand
- Li-You CHANG, Assistant Professor (Chaoyang University of Technology), PhD in Translation Studies (University College London, UK).
- Hu CHENG, Binzhou University, PhD in Education (Jeonbuk National University), Master's in Translation (Guangxi University).
- Jie LYU, PhD in Translation Studies, Lecturer (Guangdong University of Foreign Studies), expert in audiovisual translation and Business translation.
- Junchao WANG, Journal Editor, PhD in Translation Studies and Master's in Business English Studies (Guangdong University of Foreign Studies).



JHK Press, USA

General inquiry: [jackhawk@jhkpress.com](mailto:jackhawk@jhkpress.com)