

Study on the Evolution of Key Words Collocation in Report on the Work of Government Over the Past Decade

Wenguo Zhang, Lizhen Zhang

Guilin University of Technology, Guilin 10596, China

Abstract: This research employs a combination of quantitative statistics and qualitative interpretation to explore the phenomenon of thematic word collocation evolution in the English version of Reports on the Work of the Chinese Government. Specifically, the research integrates data on thematic word collocations from reports from 2015 to 2024 to construct a theoretical framework for comparative analysis and interpret the differences in word collocations between these reports and corresponding expressions in the Corpus of Contemporary American English (COCA) from memetic perspective. According to the findings, it is influenced by the norms of language expression and socio-cultural contexts. In light of this, it is necessary to optimize the English versions, enhancing the translation quality and international communication effect.

Keywords: Report on the Work of Government; Evolution of Key Words Collocation; Memetics.

1. Introduction

Since the founding of the People's Republic of China, under the leadership of the Communist Party of China, China's economic strength has continued to grow, and its overall national strength has risen steadily, attracting increasing attention from foreign scholars. Report on the Work of Government is often the most convenient channel for foreigners to understand China's comprehensive strength, as they encompass developments in China's economy, culture, politics, people's livelihood, science and technology, and other aspects. Accurate and excellent translations of Report on the Work of Government not only facilitate foreign scholars in understanding China but also help promote Chinese culture, systems, paths, theories, and more.

A search on CNKI with the keywords "English translation of Report on the Work of Government" revealed that, as of October 12, 2024, there are a total of 378 related articles. Most of these studies focus on interpreting the translations using different theoretical frameworks. For instance, Li Hongxia analyzed the strategies for translating political essays into English from the perspective of Skopos Theory, using the 2010 Government Work Report as an example. However, there have been few studies that utilize corpus-based approaches to investigate the changes in collocations of key words. In light of this, the present study examines the frequency fluctuations of verb collocations associated with the three key words "development," "reform," and "innovation." It attempts to explain the phenomenon of collocation changes using Memetics, aiming to provide a new perspective and approach for the translation studies of Report on the Work of Government.

2. Corpus Collection and Data Processing

The corpus data utilized in this study originate from the official English translations of China's Report on the Work of Government sourced from Xinhua Net, covering a span of ten years from 2015 to 2024. The Corpus of Contemporary

American English (COCA) is widely recognized in linguistic research for its extensive content coverage, abundant lexical resources, and high timeliness, serving as a reliable reflection of contemporary English usage in the United States. (Si Bingyue, Li Yitong, & Wu Meixuan, 2022) Therefore, this study selects COCA as a reference group for comparison with native English collocation habits. The focus of this study is on the use of collocations centered around key thematic words in the official English translations of China's Report on the Work of Government. Hence, the primary and core task is to identify and accurately extract these crucial thematic words and their associated collocations. The data processing process is primarily divided into the following steps:

1) Corpus Preprocessing. Initially, the official English translations of China's Report on the Work of Government from 2015 to 2024 undergo preprocessing, which includes removing unnecessary format symbols and unifying text encoding formats, laying the foundation for subsequent analysis.

2) Keyword Identification. Prior to in-depth text analysis, a crucial step is the identification of keywords based on the corpus content. This process leverages the professional text analysis tool AntConc, which is widely used in corpus linguistics research due to its powerful word frequency statistics and contextual concurrence functions. By using AntConc, high-frequency words in the reports are identified. These high-frequency words are the items that researchers need to carefully examine their collocations within the vast corpus, commonly referred to as keywords. In the context of corpus research, the concept of a keyword specifically refers to words that exhibit extremely high recurrence frequencies in a particular type of text. These words not only dominate in quantity but, more importantly, deeply reflect the stylistic and thematic characteristics of such texts, serving as the key to revealing the core content of the text. Therefore, for this study, the identification and analysis of keywords constitute the core part of the investigation.

It is worth noting that in exploring the keywords, this study did not include the usage of grammatical function words (such as prepositions, conjunctions, etc.) within the scope of

consideration. This is because, compared to content words, function words primarily serve to construct grammatical structures and express logical relationships. When used alone, they often fail to directly reflect the theme or content characteristics of the text. In contrast, content words (primarily nouns) can directly refer to specific things and properties, and their collocations tend to more effectively reveal the thematic content and contextual characteristics of the text. Therefore, this study focuses on the collocations of content words.

3) Collocation Extraction. When utilizing the Cluster Index function in AntConc, special attention was paid to the setting of the context span, which refers to the contextual range on both sides of the node word (key word). To ensure the precision and relevance of the analysis, the unit of measurement for the context span was set as word forms, and punctuation marks were explicitly excluded to focus on the collocation relationships between words. Considering that the core focus of this study is on the usage of verb-object collocations, the parameters of the search function were set to "cluster size 6/ search term position: on right," meaning that when examining a node word, we would focus on the word collocations within the six positions immediately to its left, in order to capture the verb-object combinations directly related to the node word.

4) Frequency Statistics of Collocates. After extracting the verb collocations for the three key words "development," "reform," and "innovation," we entered the crucial stage of statistics on the frequency of collocate occurrences. Initially, we utilized the KWIC (Key Word In Context) indexing function in AntConc to count the frequency of verb occurrences. Subsequently, to systematically analyze the usage frequency of these verbs, we exported the verb count data from AntConc to an EXCEL spreadsheet. In EXCEL, we used formulas to calculate the percentage of each verb among all extracted verb collocations. Furthermore, leveraging EXCEL's data visualization capabilities, we generated charts to visually present the frequency distribution of these verb collocations, providing data support for in-depth research and presentation of findings.

5) Comparative Analysis. Firstly, we entered the search pattern "Verb+Key Word" into the COCA search interface, where "Verb" represents any verb and "Key Word" was replaced with "development," "reform," and "innovation" respectively, to obtain lists of the most frequently collocated verbs with these three key words in the COCA corpus. We referenced the top 100 verbs collocated with each key word in COCA. By combining quantitative statistics with qualitative interpretation, we revealed the similarities and differences in their collocation usage and explored the cultural and linguistic factors behind the changes in collocations.

3. Memetics

3.1. Overview of Memes

In the final chapter of his bestselling book *The Selfish Gene* published in 1976, Richard Dawkins, an animalist from the University of Oxford in the United Kingdom, first proposed the unit of cultural transmission: the meme. (He Ziran & He Xuelin, 2003) Initially, memes were viewed as units of cultural transmission or imitation, encompassing various patterns in life such as melodies, ideological trends, catchphrases, fashionable clothing, house construction, and

tool manufacturing. (Dawkins, 1976) Later, memes were considered information units in the brain, replicating factors that exist in the mind and are spread through imitation. (Dawkins, 1982)

3.2. Memes and Language

Memetics centers on the competitive mechanism among memes, where memes compete with each other to invade and occupy people's minds and spread, following the principle of "survival of the fittest." Language itself is a form of meme, and words, phrases, sentences, and even texts can become linguistic memes if they are replicated and spread through imitation. In word collocations, facing immense selective pressure, lexical elements undergo intense competition to achieve survival of the fittest. The vitality of a meme depends on its recognition and frequency of transmission in use. Strong memes, which are widely popular and continuously replicated and spread, typically possess characteristics such as practicality, rationality, fashion, or authority, meeting people's communicative needs and gaining widespread recognition and use. Weak memes, on the other hand, refer to those collocations that are fleeting or exist in a small, unstable range. These memes have weak vitality and gradually disappear or are forgotten in the process of transmission. (Zhaung Meiying & He Ziran, 2010)

3.3. Meme Complex

A meme complex is a group of memes that combine with each other and can be replicated simultaneously. They are not forced together through a selection process but naturally form an integrated whole for their own survival and more efficient replication and transmission. The formation of meme complexes aids the survival of memes under specific environmental conditions, increasing their chances of survival when they combine with dominant memes in those conditions in some way. (Chen Linxia & He Ziran, 2006) Meme complexes can also undergo mutations during replication and transmission. Some meme components may be eliminated due to their incompatibility with the new environment, while new meme components may join the complex, forming new meme complexes. This process of mutation and evolution allows meme complexes to continuously adapt to new environments and social demands, maintaining their vitality and influence.

4. Collocation Changes and Data Analysis

4.1. Some Mistakes in Collocations

(1) Patchwork

Patchwork refers to the phenomenon where a translator, when attempting to construct an expression in the target language, unconsciously applies the grammatical rules and collocation habits of their native language to constrain the lexical combinations in the target language, resulting in structural variants with native language characteristics.

From a memetic perspective, the dissemination and development of language can be viewed as the replication and transmission of memes. However, there is not always a one-to-one correspondence between expressions in the target language and structures and semantics in the native language. When learners try to master the target language, they often disassemble meme elements from their native language and attempt to replicate these elements into the target language,

mistakenly believing that this disassembly and reassembly can preserve the original structure and semantics unchanged. In reality, however, this simple replication process often leads to structural and semantic mismatches, resulting in collocation errors. Take some collocations in government reports as an example. When translators try to express the concept of "innovative development," they dissect it into "innovation-driven" and "development" and attempt to combine them into "innovation-driven development" for expression. However, this combination does not conform to idiomatic expression habits in the target language. This collocation error arises precisely because the translator neglects the differences in overall structure between the native language and the target language, simply copying and pasting meme elements.

(2) Substitution

Substitution is a ubiquitous phenomenon in language learning, where a translator, when attempting to express a certain concept or idea, adopts known or familiar vocabulary to replace expressions that they have not yet mastered due to a lack of corresponding target language formulations.

From a memetic perspective, substitution can be seen as the translator partially replicating native language expressions and using these replicas to fill gaps in their target language expressions. During the translation process, translators constantly extract elements from their native language meme pool and attempt to replicate them into the target language. However, due to differences in grammatical rules, lexical collocations, and other aspects between the native and target languages, translators often encounter difficulties during the replication process. Take "promoting development" as an example. In some translations, there is a phenomenon of using "enhance" to substitute for "promote." Although this substitution expresses, to some extent, what the learner intends to express, it does not fully conform to the expression habits of the target language due to a failure to correctly extract collocations from the target language meme pool. Specifically, "promote development" is a commonly used English phrase, while "enhance development" is not a common English expression, which may confuse native speakers of the target language. This substitution behavior can reduce the accuracy and idiomaticity of the expression.

4.2. Replication Patterns of Memes in Collocation Changes

Mememes are mimicked and replicated in a manner that aligns more closely with the cultural norms of the target language, based on the inheritance and rejection of certain mememes, thereby achieving optimization (evolution) of the translation. According to memetics, "for anything to constitute a replicator, it must support a system of rules for evolution based on variation, selection, and retention." (Chen Linxia & He Ziran, 2006) Word collocations are constantly replicated and transmitted during the translation process, undergoing variation, selection, and retention. "Every meme is a replication and inheritance of previous mememes; on the other hand, it also undergoes certain variations during replication and transmission, seeking development through these variations." (He Ziran, 2005) In the context of translation, the replication patterns of mememes in collocation changes can be understood as follows:

4.2.1. Variation

Variation refers to the discrepancies that arise during the process of replication and mimicry in translation. This study

primarily focuses on diachronic variation, which entails the emergence of diverse translations over time as mememes are influenced by continually evolving linguistic norms and sociocultural contexts. Throughout their life experiences, individuals, whether consciously or unconsciously, continuously absorb and create a wide variety of mememes, and the memory system of the brain essentially constitutes a "meme complex." Furthermore, the sociocultural context stimulates the translator's subjectivity, manifesting primarily in their tendency to utilize idiomatic expressions with Chinese characteristics.

In the collocation with "reform," the phrase "deepen reform" is not an inherent expression in English. However, it has consistently occupied a prominent position in reports. This not only reflects the importance of "deepening reform" as a national policy direction but also demonstrates its widespread recognition in the global context. As China's comprehensive national strength gradually rises and its reform and opening-up continue to deepen, this expression with Chinese characteristics has gradually been accepted by global readers and has become one of the core vocabularies used to describe China's reform process. In this process, it has gradually surpassed the traditional expression "advance" and now holds a steadfast first place in relevant policy contexts, becoming one of the most representative vocabularies.

In the collocation with "innovation," the verb "drive" ranks second in the corpus, yet it exhibits a fleeting trend in the actual use of government bulletins. This may be attributed to the gradual rise of the Chinese-characteristic expression "innovation-driven," which has replaced the expression "drive innovation." This indicates that the frequency and status of various collocations are not static but constantly changing in response to social developments and variations in translator subjectivity.

4.2.2. Selection

Selection in this context refers to the competition and choice of lexical collocations. As a collective composed of numerous replicators, the language system, under the action of its selection mechanism, inevitably accompanies the emergence of certain new collocations and the gradual disappearance of old ones. This reflects the ongoing process of innovation, accumulation, and screening within the language meme pool. The replication process of lexical collocations undergoes continuous development and evolution. Firstly, different collocations compete, with some being naturally selected and outperforming others; those with higher degrees of normativity, i.e., those that more closely conform to the expressive habits of the target language, gradually become dominant mememes in the process of dissemination, surviving and continuing to replicate and spread.

The study found that among the verbs collocated with "development," the proportion of "encourage" has shown a steady increase. A search in the Corpus of Contemporary American English (COCA) revealed that the collocation "encourage development" ranks second in frequency, closely following "promote." Meanwhile, the trend for "facilitate" has emerged from non-existence and gradually stabilized, with "facilitate development" ranking fifth in frequency in COCA upon comparison. As these two meme complexes gradually align with the expressive habits of target language readers and gain recognition and widespread use, they have evolved into dominant mememes in the reports. In contrast, while the word "boost" maintained a stable frequency of use for a period, it

has gradually declined in recent years. A comparison with the top 100 verb collocations in COCA failed to retrieve "boost development," suggesting that "boost" may have certain limitations in its specific application and acceptance within policy contexts, resulting in relatively weaker replication and dissemination abilities. Similarly, the use of "strengthen" has

gradually decreased and even tended towards oblivion in recent years, primarily due to the less normative expressions such as "strengthen development," which are considered weak memes and are gradually replaced by more idiomatic and natural expressions.

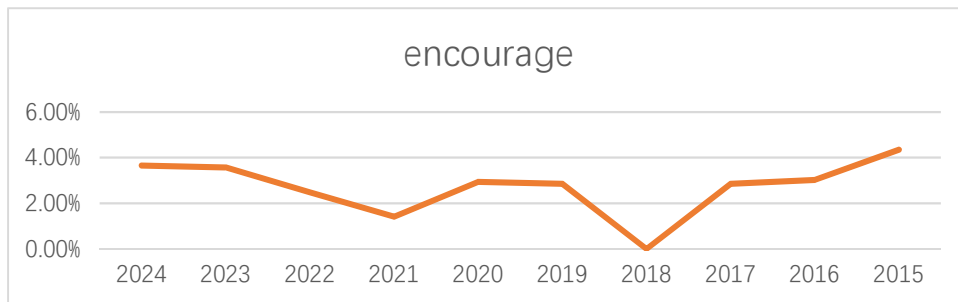


Figure 1. Trend of encourage

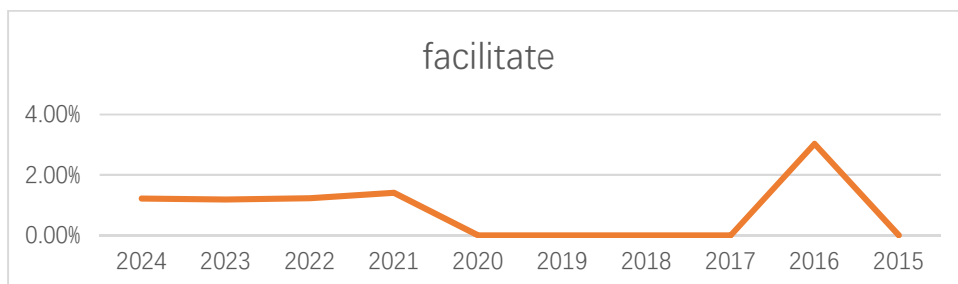


Figure 2. Trend of facilitate

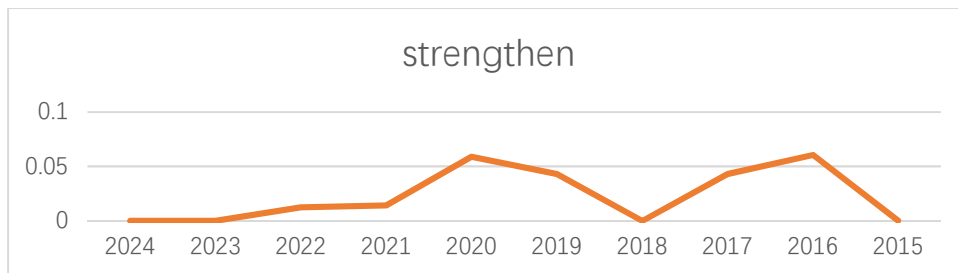


Figure 3. Trend of strengthen

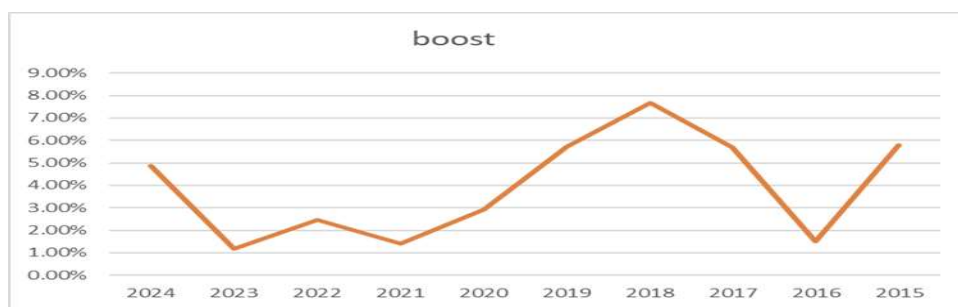


Figure 4. Trend of boost

In the collocation with "reform," the word "promote" exhibits an overall upward trend, with "promote reform" ranking second in frequency of use in the corpus. Additionally, the collocation "pursue reform" ranks seventh in the Corpus of Contemporary American English (COCA), and its frequency in reports has increased from non-existence. Due to the high replication frequency and wide dissemination of these two collocations, they have gradually stabilized in use

and become dominant memes. However, "intensify" shows a trend of gradual increase followed by eventual disappearance, and no collocation of "intensify reform" is found in the COCA corpus. This indicates that this collocation deviates somewhat from the expressive habits of the target language, affecting its acceptance and dissemination. Gradually losing its normative characteristics, it becomes a weak meme and fades away in the translation of Report on the Work of Government.

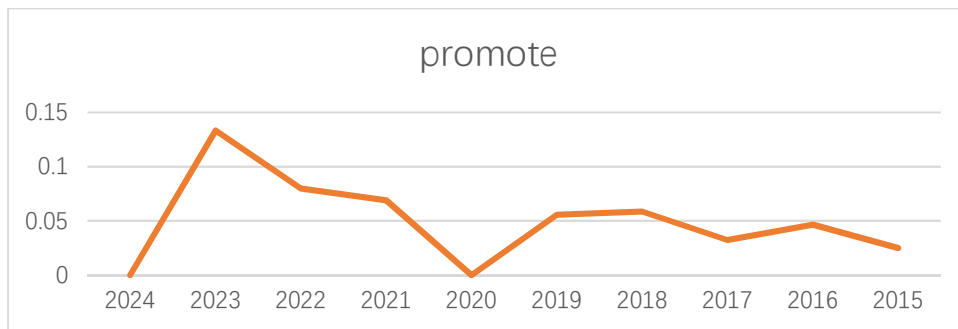


Figure 5. Trend of promote

In the collocation with "innovation," "encourage" consistently ranks first in the corpus, and its proportion has gradually increased in annual reports over time. "Support," as another verb frequently collocated with "innovation," ranks fifth in the corpus but shows a gradual upward trend in reports. Due to their alignment with the expressive norms of the target language, these two collocations have become dominant memes and have been maintained and used continuously.

However, it is noteworthy that the frequency of the verb "pursue" once surpassed "encourage" and "promote" to rank

first in the 2021 report but then gradually declined and even disappeared in recent two years. A search in the Corpus of Contemporary American English (COCA) reveals that this collocation ranks relatively low. This indicates that its "productivity" is relatively weak. People may prefer to use other phrases or vocabulary combinations, such as "spur reform" or "encourage reform," which similarly convey the active pursuit and promotion of reform and are more consistent with English expressive habits.

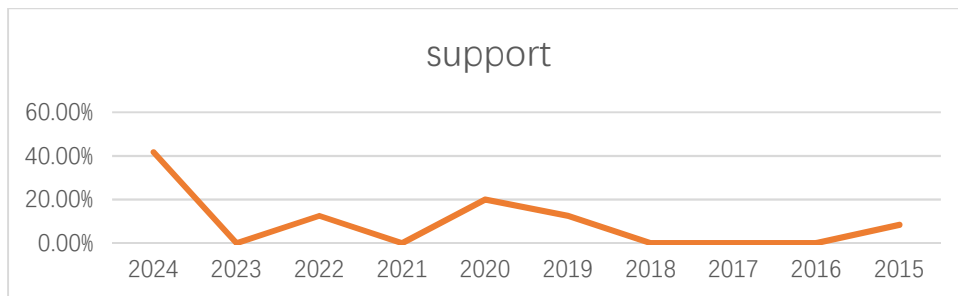


Figure 6. Trend of support

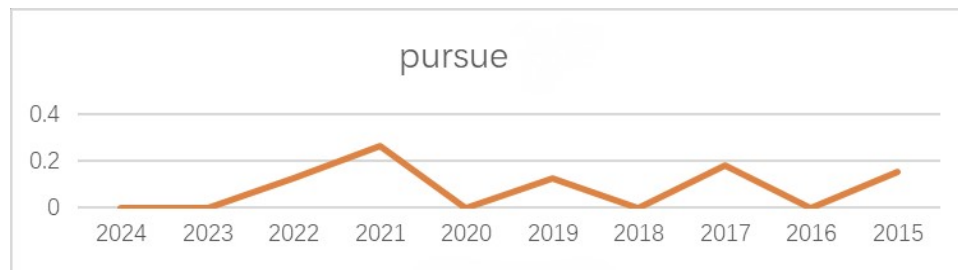


Figure 7. Trend of pursue

4.2.3. Preservation

Preservation in word collocation refers to the continuity of its usage. This continuity is based on the accuracy, normativity, and uniformity of the collocation, which are related to its acceptance within the target language culture and its degree of normalization within the target language system. The normalization of word collocations is a sign of their universal acceptance within the target language culture. Collocations with higher degrees of normativity and acceptance become dominant memes in the target language, allowing them to be preserved and transmitted over time.

In the collocation with "development," the proportions of "promote," "support," and "pursue" have remained high. In

the Corpus of Contemporary American English (COCA), among English words with similar meanings to the Chinese word "促进," "promote" ranks first, demonstrating its dominant position in expressing related concepts. "Support" follows closely, ranking third, and is also a core word in the category of verbs meaning to support or assist. Although "pursue" ranks eighth, it still holds significant influence in contexts involving the pursuit or achievement of goals. These three dominant meme combinations, due to their practicality, rationality, and authority, meet people's communicative needs and demonstrate strong vitality, enabling them to be widely popularized and continuously replicated and disseminated.

and "pursue" once led but has gradually declined. When paired with "development," "promote," "support," and "pursue" are significant, reflecting a positive attitude and subtle differences, and are widely popular due to their practicality, rationality, and authority. When paired with "reform," "advance" stands out, emphasizing positive progress, forming a fixed usage widely recognized.

In summary, the stability and variability of vocabulary collocations are influenced by various factors, including language norms, sociocultural backgrounds, and communicative needs. Dominant memes, which conform to language norms and sociocultural habits, can emerge victorious in competition and continue their replication and dissemination. Relatively weaker collocations may gradually disappear or be replaced by new ones. Regardless of the replication mode, it is crucial to adhere to the laws of meme transmission while closely aligning with the target context, fully considering the sociocultural background of the target language, language norms, and the reader's acceptance level. Only by doing so can we ensure the normalization of vocabulary collocations, achieve translation optimization and effective information transmission, thereby facilitating mutual understanding and respect among different cultures.

Acknowledgments

Project Number: JGY2024167.

Project Name: Research and Practice on the Human-Computer Collaborative Translation Education Model in MTI Programs at Science and Engineering Universities Driven by

Large Language Model Technology.

Project Number: 2022JGB208.

Project Name: Reform and Practice of English Major Construction in National First-Class Majors at Science and Engineering Universities under the Background of New Liberal Arts and Expanded Foreign Languages.

References

- [1] Dawkins, R. 1976. *The Selfish Gene* [M]. New York: OUP. (206).
- [2] Dawkins, R. 1982. *The Extended Phenotype* [M]. Oxford: OUP (109).
- [3] Chen Linxia, He Ziran. An Exploration of Linguistic Memetic Phenomena [J]. *Foreign Language Teaching and Research*, 2006, (02): 108-114+160.
- [4] He Ziran, He Xuelin. Memetics and Sociopragmatics [J]. *Modern Foreign Languages*, 2003, (02): 200-209.
- [5] He Ziran. Memes in Language [J]. *Linguistic Sciences*, 2005, (06): 54-64.
- [6] Si Bingyue, Li Yitong, Wu Meixuan. A Corpus-Based Study on the English Translation of Collocations of Generalized Verbs in Political Documents [J]. *Journal of Dalian Nationalities University*, 2022, 24(02): 168-172+177.
- [7] Zhuang Meiyong, He Ziran. Survival of the Fittest: An Analysis of the Popularity of New Words and Cool Expressions from the Perspective of Memetics [J]. *Hubei Social Sciences*, 2010, (07): 115-117.