

Literature Review of Research on English Abstracts for Scientific Papers

Chengfang Li

Southwest Petroleum University, Chengdu, China

Abstract: This paper reviews the research status and progress in the field of English abstracts of scientific papers, introduces the research background and importance of English abstracts of scientific papers as an important part of academic papers, and points out the key role of English abstracts of scientific papers in the improvement of academic communication and international influence. Through combing and analyzing the related literature, the main achievements and progresses in the research of English abstracts of scientific papers are elaborated in detail, including the analysis of the move structure, the characteristics of academic vocabulary usage, the construction of the lexicon of semantic features, and the comparison of lexical chunks, etc., and the current status and characteristics of the research in the field in China are also introduced. This paper finds that there are some hot and difficult issues in the research field of English abstracts of scientific papers, mainly in the areas of standardized writing of abstracts, cross-cultural comparison of abstract linguistic features, and the construction of automatic citation and recognition models. Finally, a comprehensive review of the existing research is conducted, pointing out the shortcomings of the current stage of research, such as the limited number of samples and the feature lexicon vocabulary to be expanded, etc., and proposing the directions and possible development trends of future research, such as the study of the writing process and cognitive mechanism, the deepening of interdisciplinary and cross-cultural contrastive research, and the application and innovation of the natural language processing technology. This paper aims to provide reference for researchers in the field of English abstracts for scientific papers, and to provide ideas and suggestions for follow-up research.

Keywords: Literature Review; English Abstract; Scientific Papers.

1. Introduction

As an important part of academic papers, the quality of writing English abstracts for scientific papers is directly related to the academic influence and dissemination effect of the papers. With the acceleration of globalization, the importance of English abstracts for scientific papers in academic communication is becoming more and more prominent. However, due to the differences in language and cultural backgrounds, non-native English speakers often face many challenges when writing English abstracts. Therefore, research on English abstracts for scientific papers not only helps to improve the writing quality of abstracts, but also promotes the international dissemination of academic achievements. Currently, research on English abstracts for scientific papers has made some progress. Researchers have made in-depth analyses of the discourse structure, linguistic features, and move patterns of abstracts from multiple perspectives. Some studies have revealed the differences between the English abstracts of scientific papers written by Chinese and foreign authors in terms of move structure and word block usage by comparing the English abstracts of scientific papers written by both Chinese and foreign authors, other studies have extracted the core semantic concepts in English abstracts of scientific papers by using corpus linguistics and constructed the corresponding semantic system. These studies provide valuable references for understanding the law of writing English abstracts for scientific papers. However, although certain results have been achieved, there are still some shortcomings in the research on English abstracts for scientific papers. For example, most of the existing studies focus on the macro-structural and linguistic features of abstracts, and do not pay enough attention to the micro-semantic features and contextual

adaptation of abstracts; in addition, there is a relative lack of research on the specificity of abstracts for different subject areas. Therefore, it is of great theoretical and practical significance to further expand and deepen the research on English abstracts for scientific papers. The purpose of this research review is to systematically sort out the current research status of English abstracts for scientific papers, analyses the existing problems and challenges, and provide directions and references for subsequent research, with a view to promoting the continuous development and progress of the field.

2. Research Method

This paper adopts the method of literature review to systematically sort out and analyses the research literature on English abstracts of scientific papers in recent years. By searching relevant databases and academic journals, 7 research papers on English abstracts of scientific papers have been collected, covering a wide range of aspects, such as the move structure, lexical features, semantic systems, and normative writing. Through in-depth reading and analysis of these literature, this paper summarizes the main achievements and views of the current research on English abstracts for scientific papers, and points out the direction of future research.

3. Findings

There are various research methods for English abstracts of scientific papers, mainly including corpus linguistics, genre analysis, multidimensional analysis and natural language processing. These methods have carried out in-depth discussions on the linguistic features, structural characteristics and semantic functions of abstracts from

different perspectives. Corpus linguistic methods conduct quantitative statistical analyses of linguistic features such as vocabulary, phrases, syntax, etc. in abstracts by constructing large corpora. For example, Cao Yan (2011) used the corpus linguistics approach to study the move distribution and usage characteristics of academic vocabulary in English abstracts of scientific journals. This method can reveal the commonalities and differences of abstract language and provide data support for abstract writing. Genre analysis focuses on the macro-structure and communicative function of abstracts. The four-step model of discourse (IMRD) proposed by Swales (1990) is the classic framework for genre analysis: introduction, method, result and discussion. Subsequent studies have expanded on this, such as Santos's (1996) five-step model of discourse (BIMRD) with the addition of the Background move. The genre analysis approach explores the communicative purpose and linguistic features of abstracts by annotating and analyzing their move structure. Multidimensional analysis methods combine quantitative and qualitative research to provide analyses of abstracts from multiple dimensions. For example, Xiao Zhonghua and Cao Yan (2014) used a combination of move analysis and multidimensional analysis methods to comparatively study the differences in move between English abstracts of scientific papers by Chinese and foreign authors. They established a multidimensional analysis model for English abstracts containing seven dimensions, and systematically examined the similarities and differences between Chinese and foreign abstracts in each dimension. Natural language processing methods use machine learning algorithms and models to automatically analyse and process abstracts. For example, Wang Lifei and Liu Xia (2017) constructed an automatic identification model of abstract move structure in English academic papers, which realized the automatic identification of abstract move structure by extracting effective move prediction features. This method improves the efficiency and accuracy of abstract analysis and provides the possibility of large-scale corpus analysis.

In recent years, domestic scholars have made significant progress in the research of English abstracts of scientific papers, mainly focusing on the analysis of move structure, characteristics of academic vocabulary usage, construction of lexicon of semantic features, and comparison of word block features. (1) move structure analysis: move structure analysis is an important aspect of research on English abstracts of scientific papers. Move is the basic unit of a discourse, which reflects the language strategy adopted by the author to achieve a specific communicative purpose. The four-step model (Introduction-Methodology-Results-Conclusion) proposed by Swales provides a theoretical basis for the move structure analysis of scientific dissertation abstracts. Domestic scholars have conducted a large number of empirical studies on this basis, which have deeply explored the differences between Chinese and foreign scientific dissertation abstracts in English in terms of language move structure. Taking the English abstracts of engineering journal papers as the research object, Hu Xin (2015) used genre analysis theory and corpus-driven phonology theory to systematically examine the similarities and differences in the structure of the word blocks used on each move and their discourse functions between Chinese authors and authors of their native languages. It was found that Chinese authors had more common missing moves on background move and used fewer word block types and fewer word blocks, which contrasted with authors of native

languages. Xiao Zhonghua and Cao Yan (2014) used multidimensional analysis methods and move structure analysis to conduct a comparative study on the differences in move structure between native language authors and Chinese authors' English abstracts in the discipline of biology, and the results showed that native language authors paid more attention to the use of intensifiers in all the mandatory move to strengthen their ideas, and demonstrated a more active participation and interactive writing style. These studies not only reveal the differences between Chinese and foreign authors' English summaries of scientific papers in terms of move structure, but also provide Chinese scholars with specific guidelines for writing English summaries that conform to international academic norms. (2) characteristics of academic vocabulary usage: academic vocabulary is an important part of English abstracts of scientific theses whose usage are directly related to the readability and academicism of the abstracts. Domestic scholars have conducted research on the use of academic vocabulary in English abstracts of scientific papers through corpus analysis and other methods. Based on a corpus of English abstracts of scientific journals, Cao Yan and Mou Aipeng (2011) examined the move distribution of academic vocabulary and its usage characteristics. It was found that the coverage rate of academic vocabulary in all move of English abstracts was more than 10%, with high frequency of use, and different collocations, class connections and semantic rhymes in different move. Some of the high-frequency word families have obvious move tendencies and special pragmatic functions. This study provides a valuable reference for grasping the law of English abstract writing from a deeper level. In addition, other scholars have explored the differences in the use of academic vocabulary in English abstracts of scientific papers in different disciplinary fields from an interdisciplinary perspective. These studies help scientists to better master the use of academic vocabulary and improve the writing quality of English abstracts. (3) construction of lexicon of semantic features: The construction of semantic feature dictionary is of great significance in realizing automatic citation and recognition of English abstracts of scientific papers in which domestic scholars have also actively explored. Song Donghuan et al. (2020) proposed a structural element BOMRC system and a method for identifying and normalizing the citation of structured abstracts based on the U.S. National Library of Medicine's structural element tagging term set and tagging classification mapping relationship. Through text mining and quantitative analysis, a semantic feature dictionary capable of structural element recognition was constructed, and validity was tested using the unstructured abstract test set. The results show that the semantic feature dictionary can effectively identify various elements of unstructured abstracts, and can be used to optimize the automatic recognition model based on machine learning methods. This study not only provides new ideas and methods for automatic citation and recognition of English abstracts of scientific papers, but also lays the foundation for intelligent processing of large-scale scientific paper abstracts. (4) comparison of word block features: Word chunks are fixed or semi-fixed multi-word combinations in language with specific semantics and functions. In English abstracts of scientific papers, the high frequency of word chunks reflects the text composition characteristics of the abstract genre. Domestic scholars reveal the important role of word chunks in abstract writing by comparing the characteristics of word

chunks in English abstracts of scientific papers in China and abroad. Hu Xin's (2015) study found that the target word blocks used by Chinese authors and native language authors at each move differed significantly in both structural form and discourse function. Native-speaking authors were more likely to use word blocks with complex structures and diverse functions to express academic views and research methods, while Chinese authors were more likely to use word blocks with simple structures and single functions. This study helps Chinese scholars to write English abstracts that comply with international academic standards from a deep semantic-functional level.

4. Discussion

Although significant progress has been made in the field of research on English abstracts for scientific papers in China, there are still some hot and difficult problems that need to be solved. These problems mainly focus on the standardized writing of English abstracts, cross-cultural comparison of abstract language features and the construction of automatic citation and recognition models. First of all, the standardized writing of abstracts is the key to improve the international influence of scientific papers. However, at present, the degree of standardization of English abstracts of scientific papers in China still needs to be improved. On the one hand, some scholars have insufficient understanding of the importance of English abstracts, which leads to irregular and low-quality abstract writing; on the other hand, there are differences in the writing requirements of English abstracts in different disciplinary fields, and there is a lack of unified norms and standards. Therefore, how to develop scientific and reasonable norms and standards for English abstract writing and improve the abstract writing ability of scientific and technological workers is one of the important hotspots in the current research field of English abstracts for scientific and technological papers. Future research should further explore the requirements and norms for writing English abstracts in depth, and put forward more targeted and operable guiding suggestions by combining the characteristics and practical needs of different disciplines. Secondly, with the increasing frequency of international academic exchanges, cross-cultural comparative research on English abstracts of scientific papers has gradually become a hot spot. When scholars from different cultures write English abstracts, they may be influenced by their own language habits and cultural traditions, resulting in differences in the linguistic features of the abstracts. However, there are still relatively few studies in this field in China, and there is a lack of systematic cross-cultural comparative studies. Future research should further expand the vision and depth of cross-cultural comparison, and explore the similarities and differences in the linguistic features of English abstracts of scientific dissertations in different cultural contexts and their causes. Through cross-cultural comparative research, it can not only enhance the understanding of academic writing styles in different cultures, but also provide useful reference and reference for the standardized writing of English abstracts of scientific papers. Finally, with the continuous development of information technology, automatic citation and recognition of English abstracts of scientific papers has become one of the research hotspots. The automatic citation and recognition model can achieve rapid processing and intelligent analysis of a large number of scientific and technological thesis abstracts, and improve the efficiency of information retrieval and use.

However, the construction of automatic citation and recognition model still faces some difficult problems. On the one hand, the linguistic features of English abstracts of scientific papers are complex and diverse, which makes it difficult to accurately cite and identify them by simple rules or algorithms; on the other hand, the differences in writing requirements and specifications for abstracts in different disciplinary fields and scientific journals have further increased the difficulty of automatic citation and recognition. Future research should further explore the automatic citation and recognition methods and techniques applicable to English abstracts of scientific papers, and improve the accuracy and robustness of the model by combining advanced technological means such as deep learning and natural language processing. At the same time, interdisciplinary cooperation and communication should be strengthened to jointly promote the development and application of automatic citation and recognition of English abstracts of scientific papers.

In view of the shortcomings of existing research, future research on English abstracts for scientific papers can be carried out in the following aspects, with a view to further deepening our understanding of abstract writing and improving the quality and international influence of abstracts. Existing studies have mostly focused on linguistic features, structural analyses, and comparative studies of abstracts, but have explored less about the cognitive process and psychological mechanisms of abstract writing. Future research can delve into the cognitive process of abstract writing, including how authors conceptualize, organize and express abstract content. Through empirical studies, the authors' attention allocation, information processing and decision-making processes when writing abstracts can be analyzed. This will help reveal the nature and laws of abstract writing and provide a scientific basis for teaching and tutoring abstract writing. For example, we can study the differences between authors with different academic backgrounds and levels of experience in the abstract writing process, and explore their different strategies in conceptualization, material selection, organization and language expression. Through comparative analyses, more effective writing methods and techniques can be identified to provide guidance for academic writing training. Although studies have been conducted to compare abstract writing in different disciplines and cultures to a certain extent, most of these studies have focused on linguistic features and have not explored the commonalities and differences in abstract writing and the reasons behind them in sufficient depth. Future research should strengthen interdisciplinary and cross-cultural comparisons, focusing not only on differences at the linguistic level, but also exploring the impact of academic norms, cultural habits, thinking patterns and other factors on abstract writing. For example, the differences in structure, vocabulary use and sentence expression in abstract writing in different disciplines (e.g., natural sciences, social sciences, humanities) can be compared, and the disciplinary characteristics and academic norms behind these differences can be analyzed. At the same time, the similarities and differences in abstract writing in different cultural contexts can be studied, and the influence of cultural factors on abstract writing style, tone and expression can be explored, which will help expand the vision and depth of abstract research and provide more comprehensive guidance and support for abstract writing in different disciplines and cultural contexts. With the rapid

development of natural language processing technology, its application in the field of text summarization is promising. Future research can use natural language processing technology to build more efficient and accurate automatic recognition and classification models for summaries, and improve the efficiency of summary processing and analysis. At the same time, new technologies such as artificial intelligence can be explored to be applied to summary writing teaching and tutoring to improve the teaching effect and students' learning experience through intelligent feedback and personalized recommendation. For example, an automatic abstract generation model can be constructed based on deep learning technology, and a large amount of abstract data can be trained so that the model can automatically generate high-quality abstracts that comply with academic norms. In addition, intelligent writing assistance tools can be developed to provide authors with real-time grammar checking, vocabulary recommendation, sentence optimization and other functions to help authors improve the quality and efficiency of abstract writing. At the same time, natural language processing technology can be used to analyses and mine a large amount of abstract data to discover the rules and trends of abstract writing and provide new perspectives and methods for abstract teaching and research.

To sum up, future research on English abstracts for scientific papers can be carried out in terms of writing process and cognitive mechanism, interdisciplinary and cross-cultural comparison, as well as the application and innovation of natural language processing technology, with a view to promoting the in-depth development of abstract writing research and improving the quality and international influence of abstracts.

5. Conclusion

As an important part of international academic communication, the quality and standardization of English abstracts of scientific papers are of great significance in enhancing the international influence of the papers. In recent years, research on English abstracts for scientific papers has made remarkable progress, but there are still some shortcomings. Future research should focus on the writing process and cognitive mechanism of abstracts, deepen interdisciplinary and cross-cultural comparative studies, and

explore the application and innovation of natural language processing technology in abstract writing. At the same time, dissertation writing courses should also strengthen the guidance and training of abstract writing, and provide targeted teaching with the characteristics of specific disciplines and cultural backgrounds, which will help improve students' ability and level of abstract writing and make greater contributions to international academic exchange.

References

- [1] Dos Santos, M. B. 1996. The textual organization of research paper abstracts in applied linguistics [J]. *Text* 16(4): 481-499.
- [2] Swales, John. *Genre Analysis: English in Academic and Research Settings*[M]. Cambridge University Press, 1990.
- [3] Cao Yan, Mou Aipeng. A study on the characteristics of lexical moves in English abstracts of scientific and technological journals [J]. *Foreign Languages Journal*,2011, (03):46-49.
- [4] Cao Yan, Zou Zhijie. Three-dimensional space of the core semantic system of English abstracts of Science and technology journals: from subject words to topic concepts [J]. *Chinese Journal of Foreign Languages*,2017,14(04):62-71.
- [5] Hu Xin. A comparative study on lexical features of English abstracts of Chinese and foreign scientific papers [J]. *Modern Foreign Languages*,2015,38(06):813-822+874.
- [6] Li Tao. Research on Standardization of English abstracts of scientific papers -- A case study of natural science papers [J]. *Journal of Liaoning University of Technology (Social Science Edition)*,2018,20(06):70-73.
- [7] Song Donghuan, Li Chenying, Liu Ziyu, et al. Construction of semantic feature dictionary for abstracts of English scientific and technological papers [J]. *Library and Information Work*, 2019,64(06):108-119. (in Chinese)
- [8] Xiao Zhonghua, Cao Yan. A comparative study of multi-dimensional steps in English abstracts of scientific papers [J]. *Foreign Language Teaching and Research*,2014,46(02):260-272+321.
- [9] Wang Lifei, Liu Xia. Construction of automatic step structure recognition model for English academic paper Abstract [J]. *Foreign Language Audio-Visual Teaching*, 2017,(02):45-50+64.