英语句法、词汇及语篇的语料库研究

沈 莹 著

英语句法、词汇及语篇的语料库研究

A Corpus-Based Study of English Syntactical Pattern, Lexicon and Contexts

沈　莹　著

英语句法、词汇及语篇的语料库研究

YINGYU JUFA CIHUI JI YUPIAN DE YULIAOKU YANJIU

沈　莹　著

Contents

Chapter One

A Corpus−Paradigm Study of English Syntactic Pattern Density and Contextural Features

1.1   Introduction

The chapter is an introduction of the research. It consists of research background, research purpose, research significance and overall organization of the research.

1.1.1   Research Background

A new language research method——corpus linguistics has sprout up gradually since the mid−to−late 20th century. Corpus linguistics, different from traditional linguistics, focuses on a great deal of real language data, conducts all−round multilevel and multidimensional research, and reveals the connotative universal rules. The appearance of corpus linguistics has exerted great influence on language research. Since 1990s, corpus linguistics has become increasingly significant. The study on corpus has developed rapidly and is still on the rise. The research on corpus and corpus linguistics has brought an innovative change to language research and language application in the last decades (Hunston, 2006). Currently, the research method based on corpus has extended gradually into many other fields, such as language teaching, translation research, lexicography, discourse analysis, especially the research on English teaching, which is a vital tendency in recent years.

Among the English integrated competences, writing competence becomes increasingly important. Besides, English writing, to a large extend, reflects an English learner's language level. Furthermore, the level of English writing can be reflected by its textual characteristics. Therefore, the study on the textual characteristics of English writing becomes a hotter and hotter topic both at home and abroad in recent years. Genres of English writing can be divided into four different

sorts: narrative, descriptive, expository and argumentative writing. Among the four different types of English writing, argumentative writing is the most important and the most difficult one to master. In addition, argumentative writing is one of the most-used styles in our work environment and daily life. Moreover, as the main composition genre of Test for English Major Band 4 (TEM-4) and Test for English Major Band 8 (TEM-8), argumentative writing is a significant part of Chinese college students' English writing (Lei Lei, 2006). Thus, the research on English argumentative writing, which can bring some vital enlightenments to the teaching of English argumentative writing, is more of practical value. Therefore, the present study, based on corpus linguistics, employing contrast analysis, makes contributions to revealing the characteristics of Chinese students' argumentative writings from different dimensions and discovering the differences between argumentative writings of the high score group and the low-score group of Chinese English learners so as to put forward some suggestions to the teaching of Chinese English argumentative writing and improve the writing level of Chinese English learners' argumentative writing.

1.1.2　Research Purpose

The present research is conducted to carry out a multidimensional analysis on Chinese English Learners' argumentative writings from the perspective of corpus linguistics. Specifically speaking, the aim of the research is to explore the following four aspects:

1) To compare the differences lying in connectives between the argumentative writings of the high-score group and the low-score group of Chinese English learners.

2) To learn the differences existing in readability between the

argumentative writings of the high-score group and the low-score group of Chinese English learners.

3) To examine the differences in syntactic density between the argumentative writings of the high-score group and the low-score group of Chinese English learners.

4) To explore the correlation between scores of argumentative writings and connective, readability and syntactic pattern density in order to provide efficient strategies for students to improve their argumentative writing level and provide beneficial suggestions for the teaching of English argumentative writing.

To sum up, the objective of the research is, first of all, to figure out respectively the characteristics of the argumentative writings of Chinese English learners in connectives, readability and syntactic density; secondly, to investigate the differences between characteristics of the argumentative writings of the high-score group and the low-score group of Chinese English learners so as to comprehend the weaknesses of the argumentative writings of the low-score group of Chinese English learners and learn from the strengths of the argumentative writings of the high-score group of Chinese English learners.

1.1.3   Research Significance

Currently, the research on the text characteristics of English argumentative writing has been a heated topic at home and abroad. Although teachers and students have devoted large amounts of time and endeavor to improving English argumentative writing skills, relevant research is still essential to provide suggestions and strategies to English argumentative writing and the teaching of it.

Previous studies on the text characteristics of English argumentative

writing both at home and abroad mainly focus on the variety of vocabulary and the complexity of syntax, but there is rare research studying English argumentative writing from the perspectives of connective, readability and syntactic pattern density by using the corpus analyzing software: Coh–Metrix. Furthermore, there is even less research making a contrast analysis between the English argumentative writings of the high–score group and the low–score group of Chinese English learners from the aforementioned three aspects. Therefore, in order to fill in the gaps, the writer intends to investigate the differences between the argumentative writings of the high–score group and the low–score group of Chinese English learners from the perspectives of connectives, readability and syntactic pattern density and explore the correlation between English argumentative writing scores and the above–mentioned three aspects so as to shed light on the teaching of English argumentative writing in the future.

1.1.4 Overall Organization

This chapter is composed of six sections.

The first section is an introduction which generally accounts for the background, the purpose, the significance and the overall organization of the research.

The second section describes the theoretical framework in which some relevant theories, such as corpus, corpus linguistics, argumentative writing, discourse analysis and so on, are presented.

The third section reviews the previous studies in which Coh–Metrix and the Spoken and Written English Corpus of Chinese Learners 2.0 (SWECCL 2.0) were used to conduct the research. Furthermore, previous research in English argumentative writings

from the perspectives of connective, readability and syntactic pattern density is also reviewed.

The fourth section deals with the research methodology, in which research questions are put forward, research materials are laid out, research instruments and research procedures are explained clearly.

The fifth section is dedicated to the results and discussions of the research where multilevel contrast is conducted. Firstly, it compares the high—score group and the low—score group of Chinese EFL learners from the perspective of connective; then contrasts the high — score group and the low—score group of Chinese English learners from the aspect of readability; thirdly, investigates the differences in the syntactic pattern density between the high—score group and the low—score group of Chinese English learners. Besides, the correlation between scores and the aforementioned three perspectives are also discussed.

The sixth section is a conclusion in which the major findings, limitations, implications and suggestions for future research are summarized.

1.2   Literature Review

This part is a review of literature related to the research at home and abroad. It retrospects the previous studies in which the text analyzing software Coh—Metrix and the Spoken and Written English Corpus of Chinese Learners 2.0 (SWECCL 2.0) are used to conduct research. Furthermore, previous studies in English argumentative writings from the perspectives of connective, readability and syntactic pattern density are also reviewed.

1.2.1 Studies by Means of Coh—Metrix

1.2.1.1 Studies by Means of Coh—Metrix at Home

Liang (2006) adopts natural language processing tool Coh—Metrix to measure the textual coherent competence of 120 students' English composition, analyzes the relationship between textual coherent competence and composition grade, and reveals the developing and changing model of Chinese students' English written language's textual coherent competence. The research manifests students' textual coherent competence and whole writing competence develop almost synchronously, but there are differences among students' textual coherent competence. And high—score composition applies global coherent devices while low—score composition relies more on local coherent devices.

This research divides textual coherence into local coherence and global coherence. Additionally, it analyzes development law of Chinese English learners' writing textual coherence from the above—mentioned two perspectives. Furthermore, it employs natural language processing tool to measure textual coherence which is more objective and makes contributions to designing automatic grading procedures of composition. Nevertheless, the research on textual coherence is limited to the study of connectives in text.

Gui (2010) employs Coh—Metrix computer evaluation and carries on a deep research on the semantics and semantic relations in the vocabulary of first language (L1) and second language (L2)'s composition to figure out the characteristics of L1 and L2 learners' employment of vocabulary in writing. The research result reveals that there are evident differences in the vocabulary level of L1 and L2's composition, which exerts certain influences on the whole quality of

composition. The major differences lie in the following: (1) By comparison with L1's composition, L2's composition relatively uses less given information to construct the co-referential relation of vocabulary to achieve the coherence of text; (2) L1's composition has more extensive and various semantic net relation than L2's composition which reflects the variety and proficiency of vocabulary employing in L1's composition; (3) L2's composition has a higher readability than L1's composition.

However, this research just makes quantitative analysis of six dependent variables reflecting L1 and L2's writing vocabulary level to examine its contained meaning and semantic relation. As for the factor causing the differences between L1 and L2's vocabulary level, it is not mentioned. Moreover, the research is only conducted in vocabulary level, while it does not make a comprehensive evaluation from the perspective of text's content, style and syntax. Therefore, the future research can be done quantitatively and qualitatively.

Du & Cai (2013) employ the text analyzing software Coh-Metrix to investigate the language characteristics influencing the writing quality of Chinese English major's argumentation and find out that readability, word fluency, coherency and the length of sentence have a high correlation with writing quality.

While the research result has some inspiration on the pedagogy of EFL argumentative writing, it only examines the language characteristics influencing the writing quality of Chinese English major's argumentative writings, which does not represent all Chinese English learners. Therefore, the following research is suggested to enlarge its research object to get a more representative consequence and efficiently improve Chinese English learners' argumentative writing.

He & Sun (2015) take the comprehensive writing performance

of college students all over China as research object and adopts the method of computer language evaluation to examine the effects of different prompt characteristics in comprehensive writing tasks on students' writing texts' language features from the surface level and the deep level. The research result indicates that different prompt characteristics will generate different effects to the performance of Chinese students' comprehensive writing task, which embodies as follows: (1) The writing prompt in different topic fields makes text present different text characteristic influential factors; (2) Diverse task illustrations lead students to adopt different argumentative models.

The existing research on the effects of prompts on writing is majorly from the perspective of independent writing task or writing score. This research makes a breakthrough, treating comprehensive writing task of college students all over China as research object. However, it also has several limitations. The research is only limited in argumentation. Besides, it conducts the research just from the dimension of topic register and tasking explanation. Future research can broaden prompt dimensions and writing genres in order to comprehensively discuss the effects of prompt on various synthetical composition task.

To sum up, while the natural language processing tool, Coh-Merix, can be applied to objectively analyze textual characteristic, decreasing subjectivity and error of human measurements, which is a time-saving, efficient and scientific method, it still has its limitations, for example, it is not able to measure text from the perspectives of context, background knowledge and rhetoric.

In addition, the aforementioned research employs Coh-Metrix to analyze the language characteristics of English writing text from

different perspectives. Nevertheless, research from dimensions of connective, readability and syntactic pattern density is relatively rare. Therefore, future research can be done on these aspects.

1.2.1.2   Studies by Means of Coh—Metrix Abroad

Crossley & McNamara (2009) examine whether first language writers of English can be distinguished from second language writers of English by lexical differences. The natural language processing tool Coh—Metrix which measures text difficulty and cohesion, and discriminant function analysis is used in the research. The results manifest that written texts of first language writers and second language writers are different in a number of aspects concerned with writers' lexical choices, such as lexical depth of knowledge, complexity and variation.

Although the research has significant implications for lexical proficiency differences between L1 and L2, it still has some limitations. First of all, the research only takes account of lexical features and ignores other variables such as part of speech and syntactic complexity which are also vital to cohesion. The future research can attempt to use many of the other 660 indices. Secondly, it only studies two languages (English and Spanish). Future research should investigate whether the lexical differences discovered here are adaptable to other languages. Thirdly, some indices chosen from Coh—Metrix cannot represent all of the lexical knowledge. Future studies should try to examine more indices to make the result more representative.

Xin Wang & Kwangsu Cho (2010) investigate two major writing genres: technical and argumentative writing, and analyze three hundred eighty—four texts written by undergraduates through a natural language processing tool named Coh—Metrix. The results