A CORPUS-BASED STUDY OF VERB-NOUN COLLOCATIONS IN THE ARGUMENTATIVE ESSAYS OF UNIVERSITY EFL LEARNERS IN CHINA

GUO SHUCAI

UNIVERSITI SAINS MALAYSIA

2017

A CORPUS-BASED STUDY OF VERB-NOUN COLLOCATIONS IN THE ARGUMENTATIVE ESSAYS OF UNIVERSITY EFL LEARNERS IN CHINA

by

GUO SHUCAI

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

July 2017

ACKNOWLEDGEMENT

I wish to acknowledge my debt of gratitude to the many people whose contribution made this work possible.

First and foremost, I would like to express my deepest gratitude to my supervisor, Associate Professor Dr. Alias Abd. Ghani, for his invaluable guidance and support in the planning and implementation of this study, for his warm-hearted and unfailing encouragement when I was faced with challenges in research and difficulties in life, for the tremendous amount of patience he had when I did not progress as expected, and for the countless hours he spent proofreading the draft copies of this work. Dr. Alias was a continual source of strength and inspiration for me on my journey of pursuing my PhD. I cannot imagine the work ever being completed without his contribution.

My sincere thanks go to Hebei University for funding my studies and providing all the support and assistance I needed to complete this work.

I am deeply indebted to members of my thesis proposal committee and examination committee for their insightful comments and suggestions; to Amelia Martin, Ymelia Stevenson, Ciara Browne, and Jimmy Redinger for the huge amount of judgment work they did for me and for the amount of time, effort, and circumspection they devoted to it; to all the students who participated in this research. Without their participation and hard work, this thesis would have been utterly impossible.

Special thanks are also extended to Dr. Zheng Zhidan and Xu Feng for assisting me with statistical data analysis; to Dr. William Barone, who gave me informative comments and valuable feedback on the first draft of my research proposal; to Dr. Kevin Lordon and Linda Lordon for all the material and emotional support they gave me throughout my doctoral studies; and to all my colleagues from Hebei University who studied at USM with me for their assistance, encouragement, and companionship during these years.

I particularly wish to convey my immense gratitude to my beloved father, who encouraged me to begin my doctoral pursuit but couldn't live to see me finish it, to my mother, my brothers and sisters for their continued support and encouragement. Most importantly, I am especially grateful to my husband and my daughter for their great sacrifice, constant support, and unshakable faith in my ability to complete my studies. Without their continuous encouragement, this work would not have been possible.

TABLE OF CONTENTS

ACKNOWLEDGEMENT……...ii

TABLE OF CONTENTS……iv

LIST OF TABLES…….....ix

LIST OF FIGURES…….....xiv

LIST OF ABBREVIATIONS……xv

ABSTRAK…….....xvii

ABSTRACT…….....xix

CHAPTER 1 – INTRODUCTION

1.0 Introduction.....1

1.1 Background to the Study.....2

1.2 Statement of the Problem……7

1.3 Research Objectives……...11

1.4 Research Questions……...12

1.5 Significance of the Study ……..13

1.6 Scope of the Study……14

1.7 Definition of Key Terms……15

CHAPTER 2 – LITERATURE REVIEW

2.0 Introduction……...18

2.1 Importance of Collocations in Language and Language Learning……...19

2.1.1 The Importance of Collocations in Native Speaker Language……19

2.1.2 The Importance of Collocations to L2 Learners……...……21

2.2 Definition of Collocations…….....23

2.2.1 Definition of Collocations in the Literature……23

2.2.2 Definition of Collocation in the Present Study……30

2.3 Classification of Collocations……...31

2.3.1 Classification of Collocations in the Literature……31

2.3.2 Classification of Collocations in This Study……35

2.4 ESL/EFL Learners' Collocational Competence……37
2.4.1 Collocational Competence of ESL/EFL Learners in Global Settings……37
2.4.2 Chinese EFL Learners' Collocational Competence……..41
2.5 Linguistic Factors and ESL/EFL Learners' Acquisition of Collocations……..47
2.5.1 Degree of Restriction of Collocations……..49
2.5.2 L1 Influence…….53
2.5.3 The Use of Delexical Verbs in Collocations…….69
2.5.4 Frequency of Collocations in Native Speaker Language……77
2.6 The Conceptual Framework of the Study……..87
2.7 Summary……90

CHAPTER 3 – RESEARCH METHODOLOGY

3.0 Introduction……...91
3.1 Research Design……91
3.2 Corpus-Based Method and the Analysis of Learner Language…….94
3.3 Samples and Sample Selection……..96
3.3.1 Sampling Technique……96
3.3.2 Setting of the Study……..97
3.3.3 Participants…….....98
3.3.4 Native Speaker Informants…….....100
3.4 Instruments ……102
3.4.1 The Essay…….....102
3.4.2 The Demographic Questionnaire……....103
3.5 References Sources……..104
3.5.1 Dictionaries……104
3.5.2 The British National Corpus……..105
3.6 Procedures and Data Analysis…….....106
3.6.1 Collecting Learner Essay Data……...107
3.6.2 Building Learner Corpus……107

3.6.3 Extracting Verb-Noun Combinations……108
3.6.3(a) Rationale: Lexemes Not Word Forms……....108
3.6.3(b) Syntactic Structures to Be Considered……...109
3.6.3(c) Verb-Noun Constructions to Be Investigated…….....110
3.6.4 Determining the Acceptability of Combinations……114
3.6.5 Distinguishing Collocations from Idioms and Free Combinations…117
3.6.6 Determining the Degree of Restriction of Collocations……119
3.6.7 Determining L1-L2 Congruency of Collocations……..127
3.6.8 Determining the Use of Delexical Verbs in Verb-Noun……130
Collocations
3.6.9 Determining the Frequency of Verb-Noun Collocations…….....133
3.6.10 Establishing Correlations Between Variables ……137
3.7 Ethical Issues……...138
3.8 Summary……139

CHAPTER 4 – RESULTS AND DISCUSSION

4.0 Introduction…….....141
4.1 Learners' Use of Verb-Noun Collocations……..141
4.1.1 Data Description…….....142
4.1.1(a) Data Overview……...142
4.1.1(b) Distribution over Essays……....142
4.1.1(c) Types and Tokens……...143
4.1.1(d) Repetition…… .147
4.1.2 Results of Acceptability Judgment…….....150
4.1.2(a) Overall Results……...150
4.1.2(b) Results for Collocations of Verbs with……..155
at Least Three Collocation Types
4.1.2(c) Results for Collocations of High Frequency……..157
4.1.3 Different Types of Deviations……158
4.1.3(a) Types of Deviant Collocations……...158

4.1.3(b) Results for Different Types of Deviations…….161
4.1.3(c) Deviations Concerning Verbs……162
4.1.3(d) Deviations Concerning Nouns……...185
4.1.3(e) Deviations Concerning the Number of Nouns……...205
4.1.3(f) Deviations Concerning Determiners……..217
4.1.3(g) Deviations Concerning Noun Complementation……...231
4.1.3(h) Deviations Concerning the Structure of Collocations…...237
4.1.3(i) Whole Collocations Inappropriate……..239
4.1.3(j) Hidden Deviations……..259
4.2 Correlation Between Degree of Restriction and Acceptability……...264
4.3 Correlation Between L1-L2 Congruency and Acceptability……...272
4.4 Correlation Between the Use of Delexical Verbs and Acceptability……...279
4.5 Correlation Between Frequency and Acceptability…….....286
4.6 Summary……...…….....292

CHAPTER 5 – CONCLUSION

5.0 Introduction…….....294
5.1 Summary of Major Findings……...294
5.1.1 Learners' Use of Verb-Noun Collocations…….294
5.1.2 Correlation Between Degree of Restriction and Acceptability……..299
5.1.3 Correlation Between L1-L2 Congruence and Acceptability……..299
5.1.4 Correlation Between the Use of Delexical Verbs and Acceptability..300
5.1.5 Correlation Between Frequency and Acceptability……....300
5.2 Implications for EFL Teaching……301
5.2.1 Raising Learners' Awareness of Collocations…….....301
5.2.2 Increasing Input Frequency of Collocations……..303
5.2.3 Selecting Appropriate Collocations to Teach…….304
5.2.4 Using Appropriate Teaching Methods…….....306
5.2.5 Improving Learners' General Vocabulary Knowledge……...309
5.3 Limitations of the Study…….....310

5.4 Directions for Future Research……..311

5.5 Summary…….314

REFERENCES……...315

APPENDICES

Appendix A: A Sample of Learners' Writing

Appendix B: The Demographic Questionnaire

Appendix C: The Consent Form

LIST OF PUBLICATIONS

LIST OF TABLES

		Page
Table 2.1	Subtypes of Grammatical Collocations and Lexical Collocations (Benson et al., 1986)	32
Table 2.2	Standardized Frequency of Collocation Errors in the Five Sub-corpora (C. Wang, 2007)	43
Table 2.3	Collocation Errors for Non-English Majors and English Majors (Y. Zhang, 2008)	44
Table 2.4	Students' CC3 Error Crosstabulation (Zhao, 2005)	63
Table 3.1	Acceptability Judgment Scheme Adapted from Nesselhauf (2005)	117
Table 3.2	Noun Collocates of *Meet* in the Sense "to satisfy (a demand, etc.)"	126
Table 4.1	Type-Token Ratio of Collocations of Top Ten Verbs With Most Collocation Types	144
Table 4.2	Collocations in at Least Ten Essays	146
Table 4.3	Collocations Repeated in the Essays of Three Learners	148
Table 4.4	Collocations That Occurred at Least Three Times in One Essay	149
Table 4.5	Overall Results of Acceptability Judgments	151
Table 4.6	Results for Collocations of Verbs With at Least Three Collocation Types	156
Table 4.7	Results for Collocations Over Ten Occurrences	157
Table 4.8	Distribution of Collocations Across Collocation Categories and Subcategories	161
Table 4.9	Distribution of Deviations Across the Deviation Types	162
Table 4.10	Distribution of Deviant Verbs Over Essays	163

Table 4.11	Distribution of Deviant Verbs Across Collocation Categories	164
Table 4.12	Verbs Produced Inappropriately in More Than Two Essays	166
Table 4.13	Verbs Confused With Other Verbs in More Than Two Essays	170
Table 4.14	Confusions Between Seven Frequent Verbs	178
Table 4.15	Verbs More Frequently Suggested as Corrections Than Appropriately Produced	179
Table 4.16	Collocations Corrected to Collocations of *Put*	182
Table 4.17	Frequencies in Three Learner Corpora of Verbs More Frequently Attempted Than Produced	183
Table 4.18	Deviant Verbs Concerning Particular Concepts	185
Table 4.19	Distribution of Deviant Nouns Across Collocation Categories	186
Table 4.20	Distribution of Deviant Nouns Over Essays	186
Table 4.21	Inappropriate Nouns in More Than Two Essays	186
Table 4.22	Distribution of Frequently Deviant Nouns Across Deviation Categories	187
Table 4.23	Deviations Regarding Types of Nouns	197
Table 4.24	Distribution of Noun-Number Deviations Across Collocation Categories	207
Table 4.25	Types of Noun-Number Deviations	208
Table 4.26	Deviations With Number of Noun out of Ignorance of Arbitrary Restrictions	209
Table 4.27	Deviations With Number of Noun out of Confusion Between Countable and Uncountable Nouns	210

Table 4.28	Frequencies of *Pollution*, *Transportation*, *Transport*, and *Work* as Countable and Uncountable Nouns in HbUC and CLEC	211
Table 4.29	*Traffic Jam* as Countable and Uncountable Nouns in Three Corpora	212
Table 4.30	Collocations With Deviations Involving the Same Noun Over Three Times	213
Table 4.31	Singular Nouns Used for Plural Nouns	215
Table 4.32	Distribution of Deviant Determiners Across Collocation Categories	219
Table 4.33	Types of Determiner Deviations	219
Table 4.34	Article Deviations Within and Across Categories	220
Table 4.35	Collocations With Frequent Determiner Deviations	221
Table 4.36	Comparison of Determiner Use in Three Corpora	224
Table 4.37	Deviations in Noun Complementation	231
Table 4.38	Inappropriate Prepositions in Noun Complementation	236
Table 4.39	Whole Collocations Inappropriate	240
Table 4.40	Existing Delexical Structures for Single Verbs	241
Table 4.41	Invented Delexical Structures for Single Verbs	244
Table 4.42	Non-Delexical Structures for Single Verbs	248
Table 4.43	Collocations for Linking Verb + Adjective/Noun Phrases	249
Table 4.44	Collocations for (Be) + Prepositional Phrases	251
Table 4.45	Collocations for Make Sb./Sth+Adjective/Noun Phrases	254
Table 4.46	Collocations for Causative Structures	256
Table 4.47	Verb-Noun Collocations for Noun-Verb Combinations	256

Table 4.48	Verb-Noun Combinations in the BNC With Similar Meaning to *Solve a Problem*	260
Table 4.49	Expressions With Similar Meaning to *Take* + *Bus*	261
Table 4.50	Verb Collocates of *Convenience* in Four Corpora	263
Table 4.51	Distribution of Collocations in Data Set 1 Across Collocation Categories	265
Table 4.52	Number of Nouns Verbs Collocate With: Examples	267
Table 4.53	Definition of Variables and Descriptive Statistics for Degree of Restriction (1)	268
Table 4.54	The Mean and Medium Variance Tests for Degree of Restriction Between the Appropriate Group and the Inappropriate Group (1)	268
Table 4.55	Probit Regression Analysis of Degree of Restriction (1)	269
Table 4.56	Definition of Variables and Descriptive Statistics for Degree of Restriction (2)	269
Table 4.57	The Mean and Medium Variance Tests for Degree of Restriction Between the Appropriate Group and Inappropriate Group (2)	270
Table 4.58	Probit Regression Analysis of Degree of Restriction (2)	270
Table 4.59	Definition of Variables and Descriptive Statistics for Congruence	273
Table 4.60	T Test and Wilcoxon Rank Sum Test for Congruence Between the Appropriate Group and the Inappropriate Group	273
Table 4.61	Probit Regression Analysis on Congruence	274
Table 4.62	Tokens and Types of Delexical Structures of Each Delexical Verb	279
Table 4.63	Definition of Variables and Descriptive Statistics for the Use of Delexical Verbs (1)	280

Table 4.64	T Test and Wilcoxon Rank Sum Test for the Use of Delexical Verbs Between the Appropriate Group and the Inappropriate Group (1)	281
Table 4.65	Probit Regression Analysis on the Use of Delexical Verbs (1)	281
Table 4.66	Definition of Variables and Descriptive Statistics for the Use of Delexical Verbs (2)	282
Table 4.67	T Test and Wilcoxon Rank Sum Test for the Use of Delexical Verbs Between the Appropriate Group and the Inappropriate Group (2)	282
Table 4.68	Probit Regression Analysis on the Use of Delexical Verbs (2)	282
Table 4.69	Frequency of Different Types of Deviations in Delexical Structures	284
Table 4.70	Frequency of Top 20 Most Frequent Collocation Types in the BNC	286
Table 4.71	Collocations With Less Than 4 Occurrences in the BNC	288
Table 4.72	Definition of Variables and Descriptive Statistics for Frequency	288
Table 4.73	The Mean and Medium Variance Tests for Frequency Between the Appropriate Group and Inappropriate Group	289
Table 4.74	Probit Regression Analysis for Frequency	290
Table 5.1	Collocational Grid Adapted from Channell (1981)	307

LIST OF FIGURES

		Page
Figure 2.1	Phraseological Categories From Howarth (1998a, p. 27)	50
Figure 2.2	Collocational Continuum (Howarth, 1998a, p. 28)	51
Figure 2.3	The Conceptual Framework of the Study	90
Figure 3.1	The Procedural Steps of the Research Design	93
Figure 3.2	Screenshot of BNC Search for Noun Collocates of the Verb Lemma *Meet*	124
Figure 3.3	Screenshot of BNC Search for the Frequency of *Commit Crime*	135
Figure 4.1	Distribution of Verb-Noun Collocations Over Essays	143
Figure 4.2	Screenshot of Search Result for *Public Traffic* in HbUC	190
Figure 4.3	Screenshot of *Relax Themselves* in HbUC	247

LIST OF ABBREVIATIONS

BBI	*The BBI Combinatory Dictionary of English*
BNC	British National Corpus
CAH	Contrastive Analysis Hypothesis
CCALED	*The Collins COBUILD Advanced Learner's English Dictionary*
CCED	*Collins COBUILD English Dictionary*
CLEC	Chinese Learner English Corpus
COCA	Corpus of Contemporary American English
EFL	English as a foreign language
ESL	English as a second language
ICLE	International Corpus of Learner English
IELTS	International English Language Testing System
L1	First language
L2	Second/foreign language
MI	Mutual information
NNS	Non-native speakers
OALD	*The Oxford Advanced Learner's Dictionary*
OCDSE	*The Oxford Collocations Dictionary for Students of English*
SECCL	Spoken English Corpus of Chinese Learners
SWECCL	Spoken and Written English Corpus of Chinese Learners
TEM-4	Test for English Majors-Band 4
TOEFL	Test of English as a Foreign Language

TTR	type-token ratio
VO	Verb + Object
VOC	Verb + Object + Complement
VOPO	Verb + Object + Particle + Object
VPO	Verb + Particle + Object
WECCL	Written English Corpus of Chinese Learners

KAJIAN TENTANG KOLOKASI KATAKERJA-KATANAMA BERASASKAN KORPUS DALAM ESEI ARGUMENTATIF PELAJAR EFL DI UNIVERSITI DI CHINA

ABSTRAK

Kepentingan kolokasi dalam pembelajaran dan penghasilan bahasa telah diakui dengan secara meluasnya. Namun kajian tentang kompetensi kolokasi dikalangan pelajar EFL tidak banyak dijalankan. Tujuan kajian ini ialah untuk mengkaji penggunaan kolokasi katakerja dan katanama dikalangan pelajar EFL dari China serta hubungannya diantara faktor kebolehterimaan penggunaan kolokasi mereka dan empat faktor linguistik termasuklah tahap restriksi kolokasi, kongruen diantara pelajar L1 dan L2, penggunaan katakerja deleksikal dalam kolokasi dan frekuensi pengunaan kolokasi oleh penutur natif. Data kajian telah dikumpulkan melalui korpus yang telah bangunkan sendiri oleh penyelidik berasaskankan 142 buah esei berbentuk argumentatif yang ditulis oleh pelajar major Bahasa Inggeris junior dan senior di sebuah universiti di China. Kesemua kombinasi katakerja-katanama dalam struktur sintaksis yang terpilih telah dikenalpasti berdasarkan kepada pertimbangan yang boleh diterima oleh penyelidik melalui korpus dan eviden dari kamus serta berdasarkan kepada intuisi empat penutur natif. Kolokasi Katakerja-katanama dibedakan daripada kombinasi katakerja-katanama bebas dan kata kiasan. Pelbagai jenis deviasi dikenalpasti serta faktor-faktor yang berkemungkinan yang menyebabkan deviasi telah dianalisis. Tahap restriksi kolokasi

diwakili oleh jumlah katanama dan katakerja yang berkolokasi dengan maksud atau makna yang tertentu. Jumlah tersebut telah diperolehi melalui data bahasa yang berdasarkan data BNC dan kamus kolokasi. Frekuensi pengunaan kolokasi pentur natif ditentukan oleh frekuensi kolokasi dalam data BNC yang diperolihi melalui kombinasi proses otomatik dan manual. Indentifikasi kesepadanan diantara L1 dan L2 dan pengunaan katakerja dileksikal adalah berdasarkan kriteria yang telah dimantapkan oleh kajian terdahulu. Dapatan kajian menunjukkan 34% kolokasi yang dihasilkan oleh pelajar adalah tidak tepat manakala hampir 40% adalah tersasar. Ini menunjukkan bahawa pelajar-pelajar yang berada pada tahap maju mengalami masalah kolokasi katakerja-katanama yang amat serius sekali. Dapatan kajian daripada analisis statistikal menunjukkan tahap korelasi restriksi kolokasi yang signifikan dengan faktor kebolehterimaan penggunaan kolokasi oleh pelajar. Lebih banyak katakerja yang berkolokasi dengan maksud yang tertentu, lebih tinggi kebarangkalian penggunaan kolokasi yang tidak tepat. Kesepadanan L1-L2, pengunaan kolokasi katakerja deleksikal dan frekuensi kolokasi menunjukkan tahap korelasi siginfikan yang positif dengan faktor kebolehterimaan. Kajian ini memberi memberi pencerahan terhadap masalah yang dihadapi oleh pelajar EFL dari China dalam penggunaan kolokasi katakerja-katanama serta mendokumentasikan pengaruh empat pembolehubah linguistik yang berkaitan dengan penggunaan kolokasi oleh pelajar berkenaan.

A CORPUS-BASED STUDY OF VERB-NOUN COLLOCATIONS IN THE ARGUMENTATIVE ESSAYS OF UNIVERSITY EFL LEARNERS IN CHINA

ABSTRACT

The importance of collocations in language learning and language production has been widely acknowledged. However, EFL learners' collocational competence and factors that may affect this aspect of their language learning remain under-investigated. The purpose of this study was to investigate Chinese EFL learners' use of verb-noun collocations and the relationship between the acceptability of their collocational use and four linguistic factors, namely, the degree of restriction of collocations, the congruence of collocations in learners' L1 and L2, the use of delexical verbs in collocations, and frequency of collocations in native speaker language. Data were collected from a self-built corpus of 142 argumentative essays written by junior and senior English majors in a university in China. All verb-noun combinations in several preselected syntactic structures were identified and subjected to acceptability judgments performed by the researcher based on corpus and dictionary evidence and by four native speakers of English based on native speaker intuition. Verb-noun collocations were distinguished from verb-noun free combinations and idioms. Different types of deviations were identified and possible causes of deviations were analysed. Degree of restriction of collocations was represented by the number of nouns the verbs collocate with in a particular sense. These numbers were obtained through exploring language data in the BNC

and a collocation dictionary. Frequency of collocations in native speaker language was represented by the frequency of collocations in the BNC, which was obtained through a combination of automatic and manual processes. Identification of L1-L2 congruence and the use of delexical verbs were based on criteria established by previous studies. The results showed that about 34% of the collocations learners produced were wrong, and close to 40% were deviant, which indicates that these advanced learners in this study are seriously deficient in verb-noun collocations. Results of statistical analyses showed that degree of restriction of collocations is significantly correlated with acceptability of learners' collocational use. The more nouns the verb collocates with in a particular sense, the higher the probability of the collocation being used inappropriately. L1-L2 congruence, the use of delexical verbs, and frequency of collocations were all found to be significantly positively correlated with acceptability. This study provides insight into the problems Chinese EFL learners have with verb-noun collocations and documents the effects of the four pertinent linguistic variables on learners' collocational use.

Chapter 1

INTRODUCTION

1.0 Introduction

Collocation is an important aspect of L2 acquisition. However, research indicates that lexical collocation is a problematic area for L2 learners, even at the advanced level (Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Howarth, 1998b; Hussein, 1990, 1998; Laufer & Waldman, 2011; Lennon 1996; Namvar & Ibrahim, 2014; Nesselhauf, 2003, 2005; Shammas, 2013). Previous studies of Chinese EFL learners' acquisition of collocations indicate that verb-noun collocational errors continue to be a major problem even for learners at advanced levels of English learning and become a major challenge for EFL teachers (L. Fan 2004; Jiang, 2006; Lin, 2006; Wang & Zhou, 2009; L. Zhang 2013; Zhao, 2005). This study aims to investigate verb-noun collocations in the argumentative essays of EFL learners at Hebei University, China. In particular, the study examines the extent to which learners' collocations conform to or deviate from target language norms and the relationship between four linguistic factors of verb-noun collocations and learners' collocation use.

This chapter will first provide a background to the study. Then it will proceed to introduce the statement of the problem, the research objectives, and research questions, followed by a discussion of the significance and scope of the study. The chapter concludes with the definition of key terms used in the study.

1.1 Background to the Study

Collocational knowledge is an important part of L2 vocabulary learning. According to Nation (2001), truly knowing a word involves knowing the nine aspects of the word: spoken form, written form, word parts, connection between form and meaning, concept and referents, associations, grammatical functions, collocations, and constraints on use. Collocation comprises the restrictions on how words can be used together. For example, one can say *powerful car*, *strong car*, *strong tea*, but not ∗*powerful tea*. Collocations like these are a pervasive phenomenon in any language, and they make up a large part of almost all types of discourse (Cowie, 1998; Hoey, 2005; Sinclair, 1991; Wray, 2002).

Collocation competence is an important indicator of language proficiency. Adult native speakers have at their disposal a large stock of collocations or other word combinations. These chunks are stored whole in a native speaker's memory and retrieved as such (Pawley & Syder, 1983). For example, *break the law*, *violate the law*, and *violate someone's privacy* are well-established collocations used by native speakers. Without such information, an ESL/EFL learner may, based on the meaning of individual words, concoct an awkward combination like *break someone's privacy*.

Many scholars have asserted that collocational competence is one important factor that contributes to the differences between native speakers and non-native speakers (Aston, 1995; Fillmore, 1979; Kjellmer, 1991; Pawley & Syder, 1983). Failure to use collocations appropriately is a major indicator of foreignness

(McArthur, 1992; McCarthy, 1990). Korosadowicz-Struzyriska (1980, p. 115, cited in Bahns, 1993, p. 56) thus described the negative effect of collocation errors, “Errors in the use of word collocations surely add to the foreign flavour in the learner's speech and writing and along with his faulty pronunciation they are the strongest markers of ‘an accent’”.

In order not to produce odd word combinations, non-native speakers need to acquire much collocational knowledge. N. C. Ellis (1997, p. 129) argued that “Speaking natively is speaking idiomatically using frequent and familiar collocations, and the job of the language learner is to learn these familiar word sequences.”

Another advantage to learning collocations is that it helps learners to develop fluency and accuracy (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Wray, 2000, 2002)). “Ready-made chunks” enable them to process and produce language at a faster rate (Conklin & Schmitt, 2008). As Lewis (1997, p. 15) pointed out, “Fluency is based on the acquisition of a large store of fixed or semi-fixed prefabricated items, which are available as the foundation for any linguistic novelty or creativity”. In addition, “It seems to be very difficult for any level of students to paraphrase or describe answers with synonymous words when they do not know the target collocations” (Koya, 2003, p.137). Therefore the importance of teaching collocations in second language pedagogy is well recognised by many other researchers (e.g., Bahns & Eldaw, 1993; Cowie, 1992; Kennedy, 2003; Lewis, 2000; Nattinger & DeCarrico, 1992).

However, collocation is constantly found to be a problematic area for ESL/EFL learners. Hussein (1990) used a multiple choice test to examine 200 junior and senior English majors' knowledge of collocations at a Jordanian university. The results showed that students' overall performance was "not satisfactory" (p. 129). Biskup (1992) investigated Polish/German advanced EFL learners' collocational competence with a translation task. She found that both groups performed poorly in the task. Bahns and Eldaw (1993) tested German advanced EFL learners' productive knowledge of verb-noun collocations with a cloze task and a translation task. The data showed that collocations presented a major problem for advanced learners' production of correct English. Another finding of the study is that learners' collocational competence did not develop at the same rate as the knowledge of vocabulary in general. Similarly, Farghal and Obiedat (1995) tested the collocational knowledge of 34 junior and senior English majors at a Jordan university and 23 English teachers who were studying for the Certification of Teachers. The former group completed a cloze task, the latter a translation task. Results indicated that both groups were seriously deficient in collocations. Subjects were found to resort heavily to strategies like synonymy, paraphrasing, avoidance, and transfer.

The findings of these studies were confirmed by more recent studies that based on larger amounts of free production data such as essays and reports. Granger (1998b) investigated the use of adverbial amplifiers (ending in *-ly*, such as *deeply*) in the ICLE (International Corpus of Learner English) and in a native speaker corpus. She found that learners underused native-like collocations and used atypical creative

combinations. Nesselhauf (2003, 2005) examined the use of verb-noun collocations by learners in the German sub-corpus of the ICLE. She found that a quarter were wrong, a third deviant (wrong or questionable). Altenberg and Granger (2001) scrutinised the collocation behaviour of the verb *make* and concluded that EFL learners, even at advanced proficiency level, have difficulty with this high frequency verb.

To sum up, collocational knowledge is an important aspect of L2 acquisition. However, it is an area of difficulty for EFL learners, even at the advanced level, regardless of their language and cultural backgrounds.

As can be seen from the above discussion, EFL learners with different L1 backgrounds have problems with collocations in general. Given the importance of collocations in language acquisition, learners' acquisition of collocations warrants due attention. In the global context, traditionally collocation has been a neglected area in linguistics and in EFL (Farghal & Obiedat, 1995). In the last twenty years, with the decline of generative influence in linguistics and the advent of the lexical approach (Lewis, 1993) in language teaching, there was a marked increase in scholarly activities on ESL/EFL learners' acquisition of collocations. Most of the earlier studies (Bahns & Eldaw, 1993; Biskup, 1992; Farghal & Obiedat, 1995; Hussin, 1990) and some of the more recent studies (Bonk, 2000; Jaén, 2007; Zughoul, 2003; Martynska, 2004) used elicitation techniques (cloze test, multiple choice test or translation tasks) to investigate learners' collocational competence and areas of difficulty. The preselected sets of collocations make it much easier to compare the

results from various subjects and groups. One limitation with this type of study is that what was tested was learners' knowledge (rather than their use) of collocations as evidenced by their judgment or introspection (Leśniewska, 2006). In addition, it is doubtful the small number of targeted collocations can represent learners' collocational knowledge. Another limitation is that these studies adopted the approach of error analysis as proposed by Corder (1971) and Richards (1970). Errors were identified, described and possible sources of errors were inferred, without taking into account non-errors. According to Harmmarberg (1974, p. 185), "This is inadequate, particularly from the language-teaching point of view." We need to know what learners do correctly as well as what they do wrongly.

In recent years, with the advent of corpus (native speaker corpus and EFL learner corpus such as the International Corpus of Learner English), modern computer technology, and complicated statistical tools and procedures, some researchers started to investigate EFL learners' collocation in written or spoken production. Collocation research with learner corpora usually adopts one of the two methodological approaches: Contrastive Interlanguage Analysis and Computer-aided Error Analysis (Granger, 2002). The first type of study (Altenberg & Granger, 2001; Durrant & Schmitt, 2009; M. Fan, 2009; Granger, 1998b; Howarth, 1998b; Juknevičienė, 2008; Laufer & Waldman, 2011; Lorenz, 1999) is contrastive and makes qualitative and quantitative comparisons between EFL learners' collocations and native speaker data. Learners are often found to overuse or underuse certain collocations compared with native speakers. The second type of study (Chi, Wong, &

Wong, 1994) focuses on EFL learners' collocation errors using computer tools to tag, retrieve and analyse them. A major strength of a corpus-based approach to learner collocation studies is that researchers can analyse a large sample of natural written or spoken language data and in a more efficient way. Corpus analysis can reveal the "hidden" aspect (overuse and underuse) of learners' collocation use, which might not be found with traditional methods of investigation (Nesselhauf, 2005, p. 41).

Although corpus-based collocation research helps to provide a more complete profile of learners' collocation use, it is largely descriptive in nature. To fully understand the nature of EFL learners' acquisition of collocations, we need to know not only what learners have or have not learned, but also what factors influence learners' acquisition. For this purpose, some researchers (Al-Zahrani, 1998; Gitsaki, 1999; Koya, 2003) investigated the relation between learners' general language proficiency and collocation competence. Some others (Shei & Pain, 2000; Web & Kagimoto, 2009) examined the effects of teaching on the learning of collocations.

1.2 Statement of the Problem

Second language acquisition is a complex process. Many factors may play a role in it (Lightbown & Spada, 1993; Spolsky, 1989). Richards and Sampson (1974) identified seven factors that have been found to influence learners' interlanguage: language transfer, intralingual interference, sociolinguistic situation, the modality of exposure to the target language and the modality of production, age, instability of the linguistic system of the learner, and the effect of the inherent difficulty of the specific items to

be learned. We can see from above that besides the various individual and contextual factors, an important factor is the inherent difficulty or "learnability" of what is to be learned. In the area of vocabulary acquisition, Laufer (1990) asked and answered the question "Why are some words more difficult than others?" Her answer to the question was that several features inherent in the word itself (pronounceability and length; part of speech, inflexional and derivational complexity; abstractness, specificity, idiomaticity; multiplicity of meaning) might affect the ease or difficulty with which the word is learnt. In the area of EFL acquisition of collocations, a similar question might be asked: Why are some collocations more difficult than others? Some linguistics factors, i.e., features inherent in the collocation itself, may have played a role.

One such factor is the similarity and differences between learners' L1 and L2. Theories about the influence of L1 on L2 acquisition contradict each other. Contrastive Analysis Hypothesis (Lado, 1957) claims that similarities facilitate acquisition, differences restrict it. Minimalist theoretical positions play down the importance of L1 (R. Ellis, 2008). A large number of studies on EFL collocation have examined the role of L1. However, most of them (e.g., Biskup, 1992; Chi et al., 1994; M. Fan, 2009; Farghal & Obiedat, 1995; Granger, 1998b; Hussein, 1990) focus on tracing the sources of learners' collocation errors or deviations to the differences between L1 and L2. Few studies have looked at both positive transfer and negative transfer by examining L1-L2 congruence and incongruence effect on learners' collocation use.

Another linguistic feature inherent in the collocation that might affect learnability is the degree of restriction of a collocation. Some words in a certain sense can only collocate with a few other words. Other words in a certain sense can collocate with a large number of words. To the best of my knowledge, only Howarth (1998b) and Nesselhauf (2003, 2005) have examined the influence of this factor. However, Howarth and Nesselhauf differed in their criteria in classifying collocations according to degree of restriction. Further research with more methodological rigor is needed to investigate this phenomenon.

The semantic characteristics of some words in collocations with certain words may also contribute to the learnability of the collocations. For example, in collocations with certain nouns, verbs like *take, make, have, do,* and *give* become delexical, in the sense that they become semantically bleached or adapted (Sinclair & Fox, 1990). In the collocation *have a rest*, the verb *have* loses its usual meaning *to possess, own* and the noun *rest* becomes the main carrier of meaning. Collocations containing delexical verbs are often regarded as problematic for EFL learners due to the fact that the meaning of the collocation is not the sum of the meaning of its constituents and that delexical verbs combine with a large number of nouns but there are also arbitrary restrictions. Some studies (e.g., Chi et al., 1994; Liao, 2010; Nesselhauf, 2005) have investigated learners' use of collocations of delexical verbs, but the results are mixed. More research is needed to examine whether the delexical nature of some words poses problems for learners.

Frequency of collocations is another factor that merits consideration. Frequency is a basic property in language (Popescu et al., 2009). Some words are more often used than others. Similarly, some collocations tend to occur more often than others in the language of a speech community. Usage-based models of language claim that language acquisition is based on one's experience with language, and frequency of exposure is a very important determining factor (Barlow & Kemmer, 2000). Since EFL learners of higher grade levels have been exposed to and have interacted with large amounts of native speaker language, it might be worthwhile to investigate whether frequency of collocation occurrence in native speaker language correlates with EFL learners' collocation use. At present, this has not been explored in depth in research on EFL learners' acquisition of collocations.

In the context of mainland China, empirical studies on EFL learners' collocations are very rare until the advent of the first large scale learner corpus CLEC (Chinese Learner English Corpus) (Gui & Yang, 2003). This is an error-tagged corpus, according to which lexical collocation errors are a major problem for learners. In particular, verb-noun collocation errors (marked as *CC3*) are by far the most common among all six types of lexical collocations, both in the corpus as a whole and in the five sub-corpora. In fact, the total number of verb-noun errors (1,542) even outnumber the errors of all other five types of lexical collocations combined (1,300).

The publication of CLEC and another learner corpus SWECCL (Spoken and Written English Corpus of Chinese Learners) motivated a large number of CLEC-based studies on learners' collocations. Most of the studies focused on

identifying the patterns of errors and possible sources of errors (e.g., Deng & Xiao, 2005; Jiang, 2006; Lin, 2006; Qin, 2005; Zhao, 2005). Some studies focused on identifying learners' overuse or underuse of collocations compared with native speaker use (Pu, 2003; Shang & Wang, 2008; Tang, 2007; Z. Wang, 2008). A few focused on examining the relationship between learners' general language proficiency level and collocation use (e.g., X. Liu 2008). Few studies have investigated the possible influence of linguistic features of collocations on learners' collocation use.

In view of the fact that Chinese EFL learners are deficient in their knowledge and competence of collocations, especially verb-noun collocations (e.g., Guo, 2003; Li, 2005; Pu, 2003; Wang & Shaw, 2008), it is worthwhile to investigate Chinese EFL learners' use of collocations, the areas of ease and difficulty for them, and the potential effects of pertinent linguistic factors on their collocational use in order to shed more light on Chinese EFL learners' acquisition of collocations and provide implications for EFL teaching and learning. The present study represents such an attempt.

1.3 Research Objectives

The primary aim of the study is to investigate Chinese EFL learners' use of verb-noun collocations in their written production, which constitutes an overwhelming proportion of the research. The secondary aim of the study is to examine the influence of four pertinent linguistic factors on learners' verb-noun

collocation use. Specifically, this study is designed to achieve the following objectives:

1. To investigate the extent to which Chinese EFL learners' use of verb-noun collocations in their argumentative essays conforms to or deviates from target language norms and particular areas of ease or difficulty for them.
2. To determine the correlation between the degree of restriction and the acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays.
3. To determine the correlation between L1-L2 congruence or incongruence and the acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays.
4. To establish the correlation between the use of delexical or non-delexical verbs and the acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays.
5. To establish the correlation between frequency in native speaker language and the degree of acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays.

1.4 Research Questions

To achieve the above research objectives, five research questions are formulated as follows:

1. To what extent does Chinese EFL learners' use of verb-noun collocations in their

argumentative essays conform to or deviate from target language norms and what are particular areas of ease or difficulty for them?

2. Is there a correlation between the degree of restriction and the acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays?
3. Is there a correlation between L1-L2 congruence or incongruence and the acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays?
4. Is there a correlation between the use of delexical or non-delexical verbs and the acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays?
5. Is there a correlation between the frequency in native speaker language and the acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays?

1.5 Significance of the Study

This study is designed to investigate all verb-noun collocations in learners' writing, instead of just focusing on erroneous collocations or a small number of preselected target collocations; hence it provides a more complete picture of learners' collocation use. In addition, the study examines the possible relations between four linguistic features and the degree of acceptability of the collocations in learners' writing. Few studies have systematically investigated the role of these factors in EFL learners' acquisition of collocations. Therefore, the results of the study add to the prior

research base and throw some new light on EFL learners' acquisition of collocations. Besides, some of the research methods used in this study (e.g., methods to determine the frequency of collocations and the degree of restriction of collocations) may provide methodological options for future researchers who explore the same issues. Finally, the pedagogical implications of the study may help the local EFL educators to better understand learners' areas of ease and difficulty regarding the learning of English verb-noun collocations, thus enabling them to make informed decisions about curriculum, material design, and teaching methods.

1.6 Scope of the Study

The scope of the research is narrowed down in four ways. Firstly, it focuses on collocations in learners' argumentative writing on a topic of general interest, since writing can reflect learners' actual language use and argumentative writing on a general topic is "fairly neutral in register and style" (Nesselhauf, 2005, p. 10), and can therefore generate the amount of writing to reveal the collocations learners often need. Secondly, among the different types of lexical collocations, the study is restricted to verb-noun collocations, because verb-noun collocations pose the greatest difficulty for Chinese learners of English, and because they "represent the propositional core of the fully-formed clause" (Howarth, 1998a, p. 163). Thirdly, the participants of the study are restricted to the junior and senior English majors at Hebei University, China. Finally, the examination of the influence of linguistics factors is restricted to the relationship between learners' collocation use and degree

of restriction of collocations, L1-L2 congruence of collocations, the use of delexical verbs, and the frequency of collocations in native speaker language use.

1.7 Definition of Key Terms

Collocations: "Combinations of two or more words whose components are used in an unidiomatic way, which follow certain structural patterns, & in which at least one word is restricted in commutability by its grammatical & semantic valency and usage" (Aisenstadt, 1981, p. 53).

Congruent collocations: Collocations that have translation equivalents in learners' native language and target language, such as *dial a number* (拨号) and *solve a problem* (解决问题) (Nesselhauf, 2005).

Degree of acceptability: The extent to which a collocation is considered to be appropriately used in context (Nesselhauf, 2005).

Degree of restriction: The level of fixedness of a collocation determined by the combinability or substitutability of its component words (Howarth, 1998b).

Delexical verbs: Verbs that are used "with nouns as their object to indicate simply that someone performs an action, not that someone affects or creates something. These verbs have very little meaning when they are used in this way" (Sinclair & Fox, 1990, p. 147).

Delexical structure: Combinations of a delexical verb and an eventive noun (Sinclair & Fox, 1990).

Error: The incorrect forms learners produce in learning a language (Gass & Selinker,

2008, p. 517).

Error analysis: "A procedure for analyzing second language data that begins with errors learners make and then attempts to explain them" (Gass & Selinker, 2008, p. 517).

Verb-noun free combinations: Verb-noun combinations in which "both the verb and the noun in the senses present in the combination can be used without arbitrary restriction on their commutability" (Nesselhauf, 2005, p. 32).

Verb-noun idioms: Verb-noun combinations in which "either no separate senses can be identified for the verb and the noun, or they can both in the given sense only be used in certain restricted environments" (Nesselhauf, 2005, p. 33).

LI transfer: "The influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired" (Odlin, 1989, p. 27).

Lemma: "the canonical form of a word" (Baker, Hardie, & McEnery, 2006, p. 104), a "set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling" (Francis & Kučera, 1982, p. 1). For example, the verb lemma TAKE consists of the words *take*, *takes*, *took*, *taking* and *taken*.

Lexeme: An abstract linguistic unit (spelt in capitals) with different variants (Jackson & Amvela, 2000, p. 56).

Negative transfer: "The use of the first language (or other languages known) in a second language context resulting in a nontarget-like second language form"

(Gass & Selinker, 2008, p. 520).

Non-congruent collocations: Collocations that do not have translation equivalents in learners' native language and target language (Nesselhauf, 2005).

Positive transfer: "The use of the first language (or other languages known) in a second language context when the resulting second language form is correct" (Gass & Selinker, 2008, p. 520).

Restricted collocations: Combinations in which commutability is arbitrarily restricted, but some commutability is possible (Howarth, 1998a).

Second language acquisition: "The learning of another language after the first language has been learned. The use of this term does not differentiate among learning situations" (Gass & Selinker, 2008, p. 521).

Verb-noun collocations: Verb-noun combinations in which "the noun can be used without arbitrary restriction in the sense in which it is being used, but the verb is, in the given sense, to some degree arbitrarily restricted to certain nouns" (Nesselhauf, 2005, p. 33).

Chapter 2

LITERATURE REVIEW

2.0 Introduction

This chapter begins with a discussion of the importance of collocations for native speakers and its importance for second language learners. Then it proceeds to introduce the different definitions of collocations adopted by two major approaches to collocation studies: the frequency-based approach and the phraseological approach. Different views of definitions and methods of analysis characteristic of each approach are described in detail. This is followed by an introduction of the different ways to classify collocations within the phraseological approach, which provides a solid basis for the present study to classify collocations. The next section reviews major research studies that have examined ESL/EFL learners' collocational competence. The following section, which constitutes a major part of the chapter, reviews the major theories and relevant research studies concerning the effects of four linguistic factors on second language learners' acquisition of collocations. The four linguistic factors are learners' native language, degree of restriction of collocations, use of delexical verbs, and frequency of collocations in native speaker language use. The chapter concludes with the conceptual framework of the present study.

2.1 Importance of Collocations in Language and Language Learning

2.1.1 The Importance of Collocations in Native Speaker Language

Collocations and other multiword units are pervasive in language. Natural language contains a large amount of recurrent multiword patterns or formulas (N. C. Ellis, 1996; 2008; N. C. Ellis, Simpson-Vlach, & Maynard, 2008; Granger & Meunier, 2008; Hill, 2000; Nattinger & DeCarrico, 1992; Lewis, 1993; Pawley & Syder, 1983; Sinclair, 1991, 2004; Wray, 2002). Pawley and Syder (1983) claimed that "The stock of lexicalised sentence stems known to the ordinary mature speaker of English amounts to hundreds of thousands" (p. 192). Hill (2000) asserted that up to 70% of the language we use or are exposed to could "be found in some form of fixed expressions" (p. 53). These claims are supported by much research evidence. Erman and Warren (2000) found that formulaic sequences of all types accounted for about 59% of their spoken language and 53% of their written language data. Biber, Johansson, Leech, Conrad, and Finegan (1999) found that the percentages of 3-word and 4-word lexical bundles were 28% in the conversation and 20% in the academic prose they studied. In an investigation to compare native speakers and learners' phraseology in academic writing, Howarth (1998b) found that the combined percentages of restricted collocations and idioms were 31% in the LOB sub-corpus and 40% in the Leeds (LUUS) corpus. Cowie (1991, 1992) reported that restricted collocations and idioms constituted 37.5% to 46% of newspaper language. This led Sinclair (1991) to propose the idiom principle: "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices,

even though they might appear to be analyzable into segments" (p. 110). He went on to propose that most normal texts are composed according to the idiom principle, whereas open-choice principle is only occasionally utilized.

The fact that multiword expressions are widespread in language can be explained by the multiple roles they play in language use. First, multiword expressions reduce speakers' language processing load and aid fluency (Barfield & Gyllstad, 2009b; Nation, 2001; Schmitt, 2004; Pawley & Syder, 1983; D. Wood, 2010b; Wray, 2002). As Pawley and Syder (1983) pointed out, a speaker may have many other things to attend to besides the syntactic structure and lexical content of his discourse. Possession of a large stock of memorised sequences simplifies the task of the speaker because ready-made chunks require little encoding work, thus allowing him to channel his energy into other activities (p. 192). In situations (like auctions, sports commentaries, and formal interviews) where high demands are placed on the real-time performance of the speaker, ready-made chunks are especially important in that they help support a difficult job (Wray, 2002). The greater the demands on working memory, the greater the need for people to rely on formulas (N. C. Ellis, 2002), because "it is easier for us to look something up than to compute it" (Bresnan, 1999). Secondly, multiword units also reduce the hearers' processing effort and supports comprehension. As Mackay (1951) stated, "Successful communication depends on symbols having significance for the receiver, and hence on their being already in some sense prefabricated for him" (p.184). Nattinger (1980) also pointed out, "For a great deal of the time anyway, language production consists