Contents

Chapter 1 Introduction **1**

1.0 Introduction **1**

1.1 Background of the Study **1**

1.2 Statement of the Problem **5**

1.3 Research Objectives **8**

1.4 Research Questions **9**

1.5 Significance of the Study **10**

1.6 Scope of the Study **10**

Chapter 2 Literature Review **12**

2.0 Introduction **12**

2.1 Importance of Collocations in Language and Language Learning **12**

2.1.1 The Importance of Collocations in Native Speaker Language 12

2.1.2 The Importance of Collocations to L2 Learners 14

2.2 Definition of Collocations **16**

2.2.1 Definition of Collocations in the Literature 16

2.2.2 Definition of Collocations in the Present Study 21

2.3 Classification of Collocations **22**

2.3.1 Classification of Collocations in the Literature 22

2.3.2 Classification of Collocations in This Study 26

2.4 ESL/EFL Learners' Collocational Competence **27**

2.4.1 Collocational Competence of ESL/EFL Learners in Global Settings 27

2.4.2 Chinese EFL Learners' Collocational Competence 29
2.5 Linguistic Factors and ESL/EFL Learners' Acquisition of Collocations 33
2.5.1 Degree of Restriction of Collocations 35
2.5.2 L1 Influence 39
2.5.3 The Use of Delexical Verbs in Collocations 47
2.5.4 Frequency of Collocations in Native Speaker Language 53
2.6 The Conceptual Framework of the Study 60
2.7 Summary 62

Chapter 3 Research Methodology 64
3.0 Introduction 64
3.1 Research Design 64
3.2 Corpus-Based Method and the Analysis of Learner Language 66
3.3 Samples and Sample Selection 68
3.3.1 Sampling Technique 68
3.3.2 Setting of the Study 69
3.3.3 Participants 70
3.3.4 Native Speaker Informants 71
3.4 Instruments 72
3.4.1 The Essay 72
3.4.2 The Demographic Questionnaire 73
3.5 Reference Sources 73
3.5.1 Dictionaries 73
3.5.2 The British National Corpus 74
3.6 Procedures and Data Analysis 75
3.6.1 Collecting Learner Essay Data 76
3.6.2 Building Learner Corpus 76
3.6.3 Extracting Verb-Noun Combinations 76
3.6.4 Determining the Acceptability of Combinations 80

3.6.5 Distinguishing Collocations from Idioms and Free Combinations 83
3.6.6 Determining the Degree of Restriction of Collocations 85
3.6.7 Determining L1-L2 Congruence of Collocations 93
3.6.8 Determining the Use of Delexical Verbs in Verb-Noun Collocations 95
3.6.9 Determining the Frequency of Verb-Noun Collocations 97
3.6.10 Establishing Correlations Between Variables 101
3.7 Ethical Issues 102
3.8 Summary 102

Chapter 4 Results and Discussion 104
4.0 Introduction 104
4.1 Learners' Use of Verb-Noun Collocations 104
4.1.1 Data Description 104
4.1.2 Results of Acceptability Judgment 111
4.1.3 Different Types of Deviations 118
4.2 Correlation Between Degree of Restriction and Acceptability 201
4.3 Correlation Between L1-L2 Congruence and Acceptability 208
4.4 Correlation Between the Use of Delexical Verbs and Acceptability 214
4.5 Correlation Between Frequency and Acceptability 221
4.6 Summary 227

Chapter 5 Conclusion 228
5.0 Introduction 228
5.1 Summary of Major Findings 228
5.1.1 Learners' Use of Verb-Noun Collocations 228
5.1.2 Correlation Between Degree of Restriction and Acceptability 229
5.1.3 Correlation Between L1-L2 Congruence and Acceptability 229

5.1.4 Correlation Between the Use of Delexical Verbs and Acceptability 230
5.1.5 Correlation Between Frequency and Acceptability 230
5.2 Implications for EFL Teaching 230
5.2.1 Raising Learners' Awareness of Collocations 231
5.2.2 Increasing Input Frequency of Collocations 232
5.2.3 Selecting Appropriate Collocations to Teach 233
5.2.4 Using Appropriate Teaching Methods 234
5.2.5 Improving Learners' General Vocabulary Knowledge 236
5.3 Limitations of the Study 237
5.4 Directions for Future Research 238
5.5 Summary 239

References 241

Appendices 266
Appendix A The Demographic Questionnaire (in Chinese) 266
Appendix B The Consent Form (in Chinese) 267

List of Tables

Table 2.1	Subtypes of Grammatical Collocations and Lexical Collocations	23
Table 2.2	Standardized Frequency of Collocation Errors in the Five Sub-Corpora	31
Table 2.3	Collocation Errors for Non-English Majors and English Majors	31
Table 2.4	Collocational Continuum	37
Table 2.5	Students' CC3 Error Crosstabulation	44
Table 3.1	Acceptability Judgment Scheme Adapted from Nesselhauf	83
Table 3.2	Noun Collocates of *Meet* in the Sense "to satisfy (a demand, etc.)"	92
Table 4.1	Type-Token Ratio of Collocations of Top Ten Verbs with Most Collocation Types	106
Table 4.2	Collocations in at Least Ten Essays	107
Table 4.3	Collocations Repeated in the Essays of Three Learners	109
Table 4.4	Collocations That Occurred at Least Three Times in One Essay	110
Table 4.5	Overall Results of Acceptability Judgments	111
Table 4.6	Results for Collocations of Verbs with at Least Three Collocation Types	115
Table 4.7	Results for Collocations over Ten Occurrences	116
Table 4.8	Distribution of Collocations Across Collocation Categories and Subcategories	120
Table 4.9	Distribution of Deviations Across the Deviation Types	121
Table 4.10	Distribution of Deviant Verbs over Essays	122
Table 4.11	Distribution of Deviant Verbs Across Collocation Categories	122
Table 4.12	Verbs Produced Inappropriately in More than Two Essays	124
Table 4.13	Verbs Confused with Other Verbs in More than Two Essays	129
Table 4.14	Confusions Between Seven Frequent Verbs	138

Table 4.15 Verbs More Frequently Suggested as Corrections than Appropriately Produced 139
Table 4.16 Collocations Corrected to Collocations of *Put* 143
Table 4.17 Frequencies in Three Learner Corpora of Verbs More Frequently Attempted than Produced 144
Table 4.18 Deviant Verbs Concerning Particular Concepts 145
Table 4.19 Distribution of Deviant Nouns Across Collocation Categories 147
Table 4.20 Distribution of Deviant Nouns over Essays 147
Table 4.21 Inappropriate Nouns in More than Two Essays 147
Table 4.22 Distribution of Frequently Deviant Nouns Across Deviation Categories 149
Table 4.23 Deviations Regarding Types of Nouns 154
Table 4.24 Distribution of Noun-Number Deviations Across Collocation Categories 160
Table 4.25 Types of Noun-Number Deviations 161
Table 4.26 Deviations with Number of Nouns out of Ignorance of Arbitrary Restrictions 162
Table 4.27 Deviations with Number of Nouns out of Confusion Between Countable and Uncountable Nouns 162
Table 4.28 Frequencies of *Pollution*, *Transportation*, *Transport*, and *Work* as Countable and Uncountable Nouns in HbUC and CLEC 164
Table 4.29 *Traffic Jam* as Countable and Uncountable Nouns in Three Corpora 164
Table 4.30 Collocations with Deviations Involving the Same Noun over Three Times 165
Table 4.31 Singular Nouns Used for Plural Nouns 167
Table 4.32 Distribution of Deviant Determiners Across Collocation Categories 169
Table 4.33 Types of Determiner Deviations 170
Table 4.34 Article Deviations Within and Across Categories 171
Table 4.35 Collocations with Frequent Determiner Deviations 172

Table 4.36 Comparison of Determiner Use in Three Corpora 174
Table 4.37 Deviations in Noun Complementation 177
Table 4.38 Inappropriate Prepositions in Noun Complementation 181
Table 4.39 Whole Collocations Inappropriate 184
Table 4.40 Existing Delexical Structures for Single Verbs 185
Table 4.41 Invented Delexical Structures for Single Verbs 188
Table 4.42 Non-Delexical Structures for Single Verbs 191
Table 4.43 Collocations for Linking Verb + Adjective/Noun Phrases 192
Table 4.44 Collocations for (*Be*) + Prepositional Phrases 194
Table 4.45 Collocations for Make Sb./Sth.+Adjective/Noun Phrases 197
Table 4.46 Collocations for Causative Structures 198
Table 4.47 Verb-Noun Collocations for Noun-Verb Combinations 199
Table 4.48 Distribution of Collocations in Data Set 1 Across Collocation Categories 202
Table 4.49 Number of Nouns Verbs Collocate with: Examples 203
Table 4.50 Definition of Variables and Descriptive Statistics for Degree of Restriction (1) 205
Table 4.51 The Mean and Medium Variance Tests for Degree of Restriction Between the Appropriate Group and the Inappropriate Group (1) 205
Table 4.52 Probit Regression Analysis of Degree of Restriction (1) 206
Table 4.53 Definition of Variables and Descriptive Statistics for Degree of Restriction (2) 206
Table 4.54 The Mean and Medium Variance Tests for Degree of Restriction Between the Appropriate Group and Inappropriate Group (2) 207
Table 4.55 Probit Regression Analysis of Degree of Restriction (2) 207
Table 4.56 Definition of Variables and Descriptive Statistics for Congruence 209
Table 4.57 T Test and Wilcoxon Rank Sum Test for Congruence Between the Appropriate Group and the Inappropriate Group 210
Table 4.58 Probit Regression Analysis on Congruence 210

Table 4.59 Tokens and Types of Delexical Structures of Each Delexical Verb 215
Table 4.60 Definition of Variables and Descriptive Statistics for the Use of Delexical Verbs (1) 216
Table 4.61 T Test and Wilcoxon Rank Sum Test for the Use of Delexical Verbs Between the Appropriate Group and the Inappropriate Group (1) 216
Table 4.62 Probit Regression Analysis on the Use of Delexical Verbs (1) 217
Table 4.63 Definition of Variables and Descriptive Statistics for the Use of Delexical Verbs (2) 217
Table 4.64 T Test and Wilcoxon Rank Sum Test for the Use of Delexical Verbs Between the Appropriate Group and the Inappropriate Group (2) 218
Table 4.65 Probit Regression Analysis on the Use of Delexical Verbs (2) 218
Table 4.66 Frequency of Different Types of Deviations in Delexical Structures 220
Table 4.67 Frequency of Top 20 Most Frequent Collocation Types in the BNC 221
Table 4.68 Collocations with Less than Four Occurrences in the BNC 223
Table 4.69 Definition of Variables and Descriptive Statistics for Frequency 224
Table 4.70 The Mean and Medium Variance Tests for Frequency Between the Appropriate Group and Inappropriate Group 224
Table 4.71 Probit Regression Analysis for Frequency 225
Table 5.1 Collocational Grid Adapted from Channell 235

List of Figures

Figure 2.1	Phraseological categories from Howarth	36
Figure 2.2	The conceptual framework of the study	62
Figure 3.1	The procedural steps of the research design	66
Figure 3.2	Screenshot of BNC search for noun collocates of the verb lemma *meet*	90
Figure 3.3	Screenshot of BNC search for the frequency of *commit crime*	99
Figure 4.1	Distribution of verb-noun collocations over essays	105
Figure 4.2	Screenshot of search result for *public traffic* in HbUC	151
Figure 4.3	Screenshot of *relax themselves* in HbUC	190

Chapter 1
Introduction

1.0 Introduction

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.

1.1 Background of 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). 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). 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.

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.

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 Hammarberg (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; 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. This indicates 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 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 linguistic 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; 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 Chinese mainland, 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; Wen et al., 2005) 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; Wang, 2008). A few focused on examining the relationship between learners' general language proficiency level and collocation use (e.g., 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 Chinese EFL learners' use of verb-noun collocations in their writing and particular areas of ease and difficulty for them.
2. To determine the relationship between the degree of restriction of verb-noun collocations in Chinese EFL learners' writing and the degree of acceptability of these collocations.

3. To examine the relationship between L1-L2 congruence or incongruence of verb-noun collocations in Chinese EFL learners' writing and the degree of acceptability of these collocations.
4. To establish the relationship between the use of delexical or non-delexical verbs in verb-noun collocations in Chinese EFL learners' writing and the degree of acceptability of these collocations.
5. To explore the relationship between the frequency of learner-produced verb-noun collocations in native speaker language and the degree of acceptability of these collocations.

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 conform to or deviate from target language norms? 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 in verb-noun collocations and the acceptability of verb-noun collocations in Chinese EFL learners' argumentative essays?
5. Is there a correlation between the frequency of the verb-noun collocations 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 to explore the same issues. Finally, the pedagogical implications of the study may help 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 linguistic 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.

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. 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 reviews the major theories and relevant research studies concerning the effects of four linguistic factors on second language learners' acquisition of collocations. 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 (Ellis, 1996, 2008; 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; 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 (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 support 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). A third role of multiword expressions for native speakers is that they help maintain the identity of individuals, and this in turn helps maintain the identity of the community (Wray, 2002).

2.1.2 The Importance of Collocations to L2 Learners

The importance of collocational knowledge to L2 learners is now widely recognised (Bahns & Eldaw, 1993; Howarth, 1998b; Pawley & Syder, 1983; Lewis, 2000; McCarthy & O'Dell, 2005; Nesselhauf, 2003, 2005). An important function of collocations is that they facilitate L2 development. There is some research evidence that children in learning a second language, as in learning L1, attend to formulaic sequences in the language input, adopt them for use and later segment and analyse them. Weinert (1995) reviewed a number of longitudinal studies which provided ample evidence that initially acquired prefabricated chunks by children were later analysed and used to facilitate overall language development. These findings were corroborated by some more recent studies (e.g., Hickey, 1993; Myles, Mitchell, & Hooper, 1999). Although research results regarding the facilitating role of formulaic language for adult L2 acquisition are mixed, N. C. Ellis (1996), in an overview of sequencing in language acquisition, concluded that memorised formulas play some important role in adult language acquisition.

Another important role of collocations and other multi-word units for L2 learners is that they make L2 learners sound native-like (Aston, 1995; Fillmore, 1979; Kjellmer, 1991; Ortaçtepe, 2013; Pawley & Syder, 1983), and failure to use them will result in foreignness (McArthur, 1992; McCarthy, 1990). As Pawley and Syder (1983) stated, native speakers do not exercise their full creative potential of syntactic rules. Among the total set of grammatical sentences, only a small proportion of them are native-like. Some expressions that are grammatically possible are not necessarily acceptable. For L2 learners, failure to conform to native norms will result in the situation described by Allerton (1984): "So often the patient language-learner is told by the native speaker that a particular sentence is perfectly good English … but that native speakers would never use it" (p. 39).

Learning collocations and other types of prefabricated language helps learners to express ideas accurately (Hill, 1999; Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Wray, 2000, 2002). As Hill (1999, p. 5) remarked, "Students with good ideas often lose marks because they do not know the four or five most important collocations of a key word that is central to what they are writing about". Hill (2000) explained that because learners don't know these collocations, they create longer utterances resulting in grammatical mistakes.

Another important role of collocations is that they help learners develop fluency (e.g., Bybee, 2002; Skehan, 1998; Wray, 2000). As pointed out by Lewis (1997, p. 15), "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". Psychological research has shown that "ready-made chunks" enable learners to process and produce language at a faster rate (Conklin & Schmitt, 2008). D. Wood (2010a) investigated L2 learners' speech fluency development (measured by some temporal variables) and their use of formulaic language over an extended period of 24 weeks. The results indicated the use of formulas was related to increased speech fluency. Therefore the importance of teaching

collocations in second language pedagogy is well recognised by many researchers (e.g., Bahns & Eldaw, 1993; Cowie, 1992; Kennedy, 2003; Lewis, 2000; Nattinger & DeCarrico, 1992).

2.2 Definition of Collocations

2.2.1 Definition of Collocations in the Literature

The term "collocation" has been used in widely different and often slightly contradictory senses in linguistics and language teaching. The only consensus among them seems to be that the term "is used to refer to some kind of syntactic relationship of words" (Nesselhauf, 2005, p. 11). Palmer and Hornby were presumably the first to use the term (Cowie, 1998). However, it is generally acknowledged that it was Firth who first brought it into linguistics as a technical term. In his paper "Modes of Meaning", Firth (1957) introduced the concept by way of examples like: "One of the meanings of *ass* is its habitual collocation with an immediately preceding *you silly*..." (p. 195). However, Firth did not give any explicit definition of the term, and "never made an attempt to expand his ideas about collocation into a theory of semantic compatibility" (Herbst, 1996, p. 380). According to Herbst (1996), the vagueness in his using the term gave rise to a number of different interpretations, resulting in the phenomenon described by Mel'cuk (1998, p. 23) "There is… no universally accepted formal definition of collocations."

Among the various uses of the term, two main views can be identified (Barfield & Gyllstad, 2009a). One of them is the "statistically oriented approach" (Herbst, 1996, p. 380) or "frequency-based approach" (Barfield & Gyllstad, 2009b, p. 3; Nesselhauf, 2005, p. 12), which sees collocations as units consisting of two or more co-occurring words within a certain distance of each other in text. This approach goes back to Firth and has since been developed further by the Neo-Firthians. A major representative of this approach is Sinclair (1991), who defined collocations as "the

occurrence of two or more words within a short space of each other in a text" (p. 170). This short space, or span, is usually set as four words on the left and four words on the right (-4, +4) of the word under study, which is called the "node". A word that co-occurs with the node in this specified environment is called a "collocate" or "collocator".

Two distinctions are often made. One is the distinction between *downward collocation* and *upward collocation*. *Downward collocation* is a collocation of a more frequent node word with a less frequent collocate. *Upward collocation* is a collocation of a less frequent node with a more frequent collocate (Sinclair, 1991, pp. 115–116). Another distinction is between "significant" collocations and "casual collocations", sometimes the former is reserved for collocations. Sinclair (2004) defined significant collocation as "the regular collocation between two items, such that they co-occur more often than their respective frequencies, and the length of text in which they appear, would predict" (p. 10). A non-significant collocation is a casual collocation. Some statistical measures have been used to identify word pairs that appear with greater than random probability. All the measures are based on the principle of comparing the number of times two words co-occur in a corpus with the number of times it would be predicted to appear by chance on the basis of their respective frequency in the corpus. The commonly used association measures are t-score and mutual information (MI) (Durrant & Schmitt, 2009). The difference between the two measures is that t-score tends to highlight very frequent collocations, whereas MI tends to emphasise collocations that are less frequent but whose constituent words are not often found apart (Stubbs, 1995).

The other view is the "phraseological approach" (Barfield & Gyllstad, 2009a, p. 5; Nesselhauf, 2005, p. 12), which is heavily influenced by the work of Russian phraseologists in the late 1940s to the 1960s (Cowie, 1998). In contrast to the frequency-based approach which uses frequency as the sole criterion to define collocations, those working in the phraseological tradition define collocations according to the semantic characteristics of

combinations and commutability of the individual elements. They treat collocations as word combinations of various degrees of fixedness in grammatical constructions. Their major focus is on distinguishing collocations from other types of word combinations and classifying collocations. Cowie, Howarth, and Nesselhauf are among the major representatives of this research tradition.

Although there is much common ground between them, their definitions of collocations vary to some extent. Cowie (1988) made a distinction between "formulae" (e.g., *good morning*, *how are you*) and "composites" (*kick one's heels, pass the buck*) according to "the kinds of meaning which their members convey and to the structural level at which they operate" (p. 132). Combinations in the group of "composites" were further divided into four categories according to the criterion of transparency (whether the elements and the combination have a literal or non-literal meaning) and the criterion of commutability (whether and to what degree the substitution of the elements is restricted). The resulting categories are thought to form a kind of continuum with no clear cut-off points between them (Nesselhauf, 2005, pp. 14–15).

Free combinations (e.g., *drink tea*)
- the restriction on the substitution can be specified on semantic grounds
- the elements of the word combination are used in the literal sense

Restricted collocations (e.g., *perform a task*)
- some substitution is possible, but there are arbitrary limitations on substitution
- at least one element has a non-literal meaning, and at least one element is used in its literal sense; the whole combination is transparent

Figurative idioms (e.g., *do a U-turn*, in the sense of "completely

change one's policy or behaviour")
- substitution of the elements is seldom possible
- the combination has a figurative meaning, but preserves a current literal interpretation

Pure idioms (e.g., *blow the gaff*)
- substitution of the elements is impossible
- the combination has a figurative meaning and does not preserve a current literal interpretation

As pointed out by Nesselhauf (2005), the major drawback of this definition for restricted collocations is that the two defining criteria (opacity and commutability), though often assumed to coincide, do not regularly coincide. There are combinations that meet the opacity criterion but not the commutability criterion. For example, combinations with *face*, such as *face a task/ a financial crisis/ her anger* would be classified as collocations according to the opacity criterion (*face* used in the figurative sense "to have to deal with a particular situation"), but free combinations according to the commutability criterion (the choice of object seems unlimited as long as it refers to some difficult or unpleasant situation). There are also combinations that meet the commutability criterion, but not the opacity criterion. In combinations like *commit a crime*, both *commit* and *crime* are used in the literal sense, but there is arbitrary restriction (? *commit a lie/ a deceit*). Besides, there are cases where both elements are used in the figurative sense (e.g., *take steps*), which would be classified as idioms based on the opacity criterion, but where the criterion of commutability is not met (*take actions*, *take measures*). To solve the problem, Nesselhauf (2005) proposed to use only one criterion, the criterion of commutability, to define collocations.

According to Nesselhauf (2005), a second problem with the definition is the vague meaning of the criterion of commutability itself, which leads to different interpretations. She illustrated this point with examples from

Aisenstadt and Cowie. Aisenstadt assumed the commutability of *shrug* is restricted to *shouders sth. off*/*away*, and the commutability of *shoulders* is restricted to *shrug*/*square*/*haunch*. It can be seen that the three verbs are not synonymous; they are just verbs that can be used with *shoulders*. In that case, *shoulders* are not restricted to the three verbs. There are a lot of verbs that can take the place of *shrug*, such as *rub*/*straighten*/*wash one's shoulders*. On the other hand, Cowie seemed to limit commutability to synonyms or near synonyms (*have*/*exert*/*exercise influence*, *pay*/*devote*/*give attention*/*heed*, *break one's journey*, **trip*, **voyage*). However, Cowie did not apply this criterion in identifying free combinations.

Nesselhauf (2005) pointed out that the fundamental problem with these attempts to distinguish collocations from other types of combinations is that they assume the lexical elements in a combination have the same status. This is not the case, however. Nesselhauf cited Mel'cuk's (1998) theory that one of the elements in a collocation is semantically autonomous, the other is not. In verb-noun collocations, it is the noun that is semantically autonomous. Nesselhauf went on to suggest that the noun in a verb-noun collocation and the noun in a free combination have the same status and therefore should not be used as a criterion to distinguish collocations from free combinations. Instead, the combinability of verbs should be used as the criterion. Based on a detailed analysis, Nesselhauf (2005, p. 30) divided verbs into five groups according to their combinatory possibilities:

1. verb combinable with (virtually) every noun (e.g., *want sb./sth.*)
2. verb combinable with a large group of nouns (e.g., *kill*+ [+ALIVE])
3. verb combinable with a small but well-delimitable semantic group of nouns (e.g., *catch* +[infectious disease])
4. verb combinable with a sizable group of nouns, but there are exceptions (e.g., *commit* +[something wrong or illegal], but ? *commit a lie*/*deceit*/*delinquency*)