

**INVESTIGATING VERB-NOUN LEXICAL COLLOCATIONAL ERRORS IN A  
MALAYSIAN ENGLISH LEARNER CORPUS**

**By**

**ANG LENG HONG**

**Thesis submitted in partial fulfilment of the requirements for  
the degree of  
Masters of Arts in Linguistics and English Language Studies**

**UNIVERSITI SAINS MALAYSIA  
2010**

## ACKNOWLEDGEMENTS

I would like to express my profound gratitude to Associate Professor Dr. Hajar Abdul Rahim, my supervisor, for her guidance and dedication in this thesis. She carefully read every draft of my chapters and gave me valuable suggestions and insightful comments throughout the preparation of this thesis. Without her, I would never have been able to complete this work.

I would like to express my gratefulness to Universiti Sains Malaysia for granting me a scholarship and providing me with everything that facilitated my study.

I would also like to thank Associate Professor Dr. Arshad Abd. Samad, the deputy dean (student and alumni affairs) of Faculty of Educational Studies, Universiti Putra Malaysia for granting me the permission to use the EMAS (The English of Malaysian School Students) corpus in my study.

I record my deepest appreciation to my family for their understanding and unfailing support. Special thanks are due to my mother, who provided me with all sorts of support throughout the entire process of my graduate study. Because of her unconditional love and warm encouragement, I have the chance to pursue my goal.

Last but not least, I thank the staff members of the School of Humanities, Universiti Sains Malaysia for their support.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS	Page ii
TABLE OF CONTENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
ABSTRAK	xi
ABSTRACT	xiv

### CHAPTER 1: INTRODUCTION

1.1 Introduction	1
1.2 Background to the study	1
1.3 Statement of the problem	5
1.4 Objectives of the study	6
1.5 Research questions	6
1.6 Significance of the study	7
1.7 Scope of the study	7
1.8 Theoretical framework of the study	8
1.9 Definitions of terms	8
1.10 Summary	10

### CHAPTER 2: REVIEW OF RELATED LITERATURE

2.1 Introduction	11
2.2 The notion of collocation	11
2.2.1 Importance of collocations in second language acquisition	15
2.3 The notion of Interlanguage (IL)	17
2.3.1 Interlanguage Hypothesis by Selinker (1974)	18
2.3.2 Characteristics of Interlanguage by Adjemian (1976)	20
2.4 The notion of error	20
2.5 Corpus-based study	22
2.5.1 Corpus-based study of learner language	22
2.5.2 Corpus-based study of collocations	23

2.6 Classification of collocations	24
2.6.1 Lexical and grammatical collocations	24
2.7 Empirical studies on collocations	25
2.7.1 Studies on the knowledge of collocations	26
2.7.2 Studies on the relation between collocations and language proficiency	30
2.7.3 Studies on the collocational error analysis	32
2.8 Error analysis in learner language	37
2.8.1 Limitations of EA	39
2.9 Types of errors	43
2.10 Sources of errors	45
2.10.1 Language transfer/ Interlingual transfer	46
2.10.2 Intralingual transfer	47
2.10.3 Communication strategies	48
2.11 Incorporation of various sources of errors into the study of collocational errors	50
2.12 The conceptual framework for the present study	52
2.13 Summary	54

## **CHAPTER 3: METHODOLOGY**

3.1 Introduction	55
3.1.1 Research questions	55
3.2 Research design	55
3.3 The corpus	56
3.3.1 The syntactic patterns considered	58
3.4 Data generation procedure	59
3.4.1 Data generation	61
3.4.2 Identification of errors	61
3.4.3 Classification of errors	63
3.4.3.1 Classification of verb errors	64
3.4.3.2 Classification of noun errors	65



3.4.3.3 Usage 1 errors	66
3.4.3.4 Usage 2 errors	66
3.4.3.5 Classification of preposition errors	67
3.4.3.6 Classification of determiner errors	68
3.4.3.7 Classification of number errors	68
3.4.4 Quantification of errors	69
3.4.5 Analysis of sources of collocational errors	69
3.4.5.1 L1 transliteration	70
3.4.5.2 Language switch	70
3.4.5.3 False concept hypothesised	71
3.4.5.4 Overgeneralisation	71
3.4.5.5 Ignorance of rule restrictions	71
3.4.5.6 Approximation	72
3.5 Summary	73

## **CHAPTER 4: DATA ANALYSIS AND FINDINGS**

4.1 Introduction	74
4.2 The overall results of types of collocational errors	74
4.2.1 Verb errors	76
4.2.2 Noun errors	78
4.2.3 Usage 1 errors	80
4.2.4 Usage 2 errors	81
4.2.5 Preposition errors	82
4.2.6 Determiner errors	85
4.2.7 Number errors	87
4.3 Overall results of sources of collocational errors	88
4.3.1 L1 transliteration/ L1 literal translation	89
4.3.2 Language switch	92
4.3.3 False concept hypothesised	93
4.3.4 Overgeneralisation	95
4.3.5 Ignorance of rule restrictions	96

4.3.6 Approximation	102
4.4 Summary	106
<b>CHAPTER 5: SUMMARY, CONCLUSION, PEDAGOGICAL IMPLICATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH</b>	
5.1 Introduction	107
5.2 An overview of the present study	107
5.3 Summary of the findings	109
5.3.1 Research question 1: What types of verb-noun lexical collocational errors are found in Malaysian learners' vocabulary?	109
5.3.2 Research question 2: What sources of verb-noun lexical collocational errors are found in Malaysian learners' vocabulary?	112
5.4 Conclusion	115
5.5 Pedagogical implications	115
5.5.1 Teaching L2 collocations	115
5.5.2 Introducing corpora and concordancer	116
5.5.3 Using dictionaries	117
5.6 Recommendations for future research	118
References	120
Appendices	
Appendix A: Concordance for erroneous verb-noun lexical collocations based on verb as node	128
Appendix B: Concordance for erroneous verb-noun lexical collocations categorised according to types of collocational errors	137
Appendix C: The distribution of sources of collocational errors and types of collocational errors in erroneous verb-noun lexical collocations	148

## LIST OF TABLES

	Page
Table 2.1 Collocational continuum (adopted from Howarth, 1998: 28)	14
Table 2.2 Categories of studies on collocations	25
Table 2.3 Procedures for error analysis (Corder, 1974b)	38
Table 2.4 Procedures for error analysis (Gass and Selinker, 2008: 103)	38
Table 2.5 Framework for classification of types of errors in collocations (Nesselhauf, 2003: 232)	44
Table 2.6 A typology of communication strategies developed by Tarone (1981: 286-287)	49
Table 2.7 Framework for classification of sources of collocational errors incorporated by Liu (1999) (adapted from Richards (1974) and Tarone, 1981)	51
Table 3.1 Syntactic patterns of verbs for verb-noun lexical collocations studied in the present study (adapted from Hunston and Francis, 2000: 51-53)	58
Table 3.2 Procedures for error analysis (Gass and Selinker, 2008)	60
Table 3.3 Procedures for error analysis in the present study (Adapted from Gass and Selinker, 2008)	60
Table 3.4 Framework for classifying types of errors in collocations in the present study (adapted from Nesselhauf, 2003: 232)	64
Table 3.5 Classification of sources of collocational errors in the present study (adapted from Richards, 1974 and Tarone, 1981)	70
Table 4.1 Types of collocational errors found in the corpus	75
Table 4.2 Collocational errors concerning verbs	76
Table 4.3 Collocational errors concerning nouns	78
Table 4.4 Collocational errors concerning prepositions	82
Table 4.5 Collocational errors concerning articles	85
Table 4.6 Collocational errors concerning number	87
Table 4.7 Sources of collocational errors in collocations	89

Table 4.8 L1 transliteration on the different types of collocational errors	90
Table 4.9 False concept hypothesised on the different types of collocational errors	94
Table 4.10 Ignorance of rule restrictions on various types of collocational errors	98
Table 4.11 Approximation on various types of collocational errors	104



## LIST OF FIGURES

	Page
Figure 2.1: The conceptual framework for the present study	52

## LIST OF ABBREVIATIONS

BNC	British National Corpus
EA	Error Analysis
EFL	English as a foreign language
ESL	English as a second language
IL	Interlanguage
L1	First Language
L2	Second Language
NL	Native Language
SLA	Second Language Acquisition
TL	Target Language

## **Kajian Mengenai Kesalahan Kolokasi Leksikal Kata Kerja-Kata Nama dalam sebuah Korpus Bahasa Inggeris Pelajar Malaysia**

### **ABSTRAK**

Kajian ini menggunakan cara mengakses data korpus berkomputer untuk mengkaji jenis dan sumber kesalahan kolokasi leksikal kata kerja-kata nama yang terdapat dalam sebuah sub-korpus kepada korpus bahasa Inggeris pelajar Malaysia, EMAS (The English of Malaysian School Students). Korpus yang terlibat mengandungi 130 karangan yang dihasilkan oleh pelajar berbangsa Melayu Tingkatan Empat yang berasal dari tiga negeri di Semenanjung Malaysia.

Kajian ini menggunakan definisi Howarth (1998) yang mengatakan bahawa kolokasi merupakan kombinasi-kombinasi perkataan termasuk kombinasi bebas dan kolokasi terhad. Walaupun terdapat dua kategori kolokasi, iaitu kolokasi leksikal dan kolokasi gramatikal (Benson, Benson and Ilson, 1986), kajian ini hanya mengkaji satu jenis kolokasi leksikal, iaitu kolokasi leksikal kata kerja-kata nama. Akan tetapi, berdasarkan rangka Nesselhauf (2003) mengenai klasifikasi jenis kesalahan kolokasi, kolokasi leksikal kata kerja-kata nama dalam kajian ini mengambil kira elemen gramatikal yang wujud dalam kolokasi-kolokasi, seperti kata sendi nama. Oleh yang demikian, definisi kolokasi leksikal kata kerja-kata nama dalam kajian ini dimodifikasikan kepada kombinasi antara satu kata kerja dan satu kata nama yang mengambil kira elemen gramatikal yang lain seperti kata sendi nama.

Kolokasi adalah mustahak kerana ia membantu meningkatkan kemahiran bahasa pelajar-pelajar dan membawa mereka ke arah kefasihan seperti penutur jati bahasa Inggeris. Hasil kajian lalu menunjukkan bahawa penutur bukan jati mempunyai ilmu pengetahuan kolokasi yang terhad berbanding dengan penutur jati. Perpindahan antara elemen-elemen dalam bahasa kedua merupakan faktor utama yang menjejaskan penghasilan kolokasi yang sesuai oleh penutur bukan jati. Selain itu, kajian-kajian lalu juga mendapati bahawa kolokasi memainkan peranan yang penting dalam menentukan tahap kemahiran bahasa penutur-penutur bukan jati.

Kajian ini dijalankan dengan berlandaskan teori Interlanguage (Selinker, 1974). Identifikasi sumber-sumber kesalahan kolokasi berlandaskan Hipotesis Interlanguage mengesahkan bahawa struktur and item linguistik dalam bahasa pelajar boleh menjadi fosil. Kajian ini menggunakan rangka Analisis Kesalahan yang dicadangkan oleh Gass dan Selinker (2008) untuk menjalankan analisis. Perisian komputer, *Wordsmith Tools* digunakan untuk menghasilkan data untuk kajian ini. Kajian ini merujuk kepada *Kamus Kolokasi Oxford* dan Korpus Nasional British untuk menentukan kolokasi yang salah. Rangka Nesselhauf (2003) digunakan untuk mengklasifikasikan dan menghuraikan pelbagai jenis kesalahan kolokasi. Klasifikasi sumber kesalahan adalah berdasarkan rangka Richards (1974) dan Tarone (1981), iaitu perpindahan elemen-elemen dalam bahasa pertama kepada bahasa kedua, perpindahan antara elemen-elemen dalam bahasa kedua dan penjelasan maksud dengan perkataan lain. Perpindahan elemen-elemen dalam bahasa pertama kepada bahasa kedua dibahagikan kepada dua sub-kategori, iaitu transliterasi dari bahasa pertama dan penggunaan bahasa pertama dalam bahasa kedua



tanpa terjemahan. Perpindahan antara elemen-elemen dalam bahasa kedua diklasifikasikan kepada tiga sub-kategori, yaitu konsep palsu dihipotesiskan, generalisasi keterlebihan dan ketidakpatuhan kepada pengehadan peraturan. Selain itu, salah satu sub-komponen di bawah kategori penjelasan maksud dengan perkataan lain, yaitu penganggaran dirujuk sebagai salah satu sumber kesalahan kolokasi dalam kajian ini. Hasil kajian ini menunjukkan bahawa antara tujuh jenis kesalahan kolokasi, jenis kesalahan yang paling banyak ditemui ialah kesalahan kata sendi nama. Bagi sumber kesalahan kolokasi, didapati bahawa perpindahan antara elemen-elemen dalam bahasa kedua merupakan sumber kesalahan kolokasi yang paling ketara antara ketiga-tiga kategori sumber kesalahan kolokasi.

## Investigating Verb-Noun Lexical Collocational Errors in a Malaysian English Learner Corpus

### ABSTRACT

This study employed a corpus-based method to investigate types and sources of verb-noun lexical collocational errors in a subcorpus of a Malaysian learner corpus, EMAS (The English of Malaysian School Students). The corpus consists of a total of 130 essays written by Form Four Malay learners from three different states in peninsula Malaysia.

The study adopted Howarth's (1998) definition which states that collocations are word combinations which include free combinations and restricted collocations. Although there are two categories of collocations, namely lexical and grammatical collocations (Benson, Benson and Ilson, 1986), this study only investigated one type of lexical collocation, which is verb-noun lexical collocations. Nevertheless, following Nesselhauf's framework for classification of types of collocational errors, the investigation of verb-noun lexical collocations in this study includes grammatical elements which occur in the collocations, such as prepositions. The term *verb-noun lexical collocations* in this study was redefined as combinations of a verb and a noun which take into account grammatical elements such as prepositions.

Collocations are important as they help to increase learners' language proficiency and bring them towards native-like fluency. The findings of the previous studies indicate that non-native speakers have a rather limited knowledge of collocations compared with native speakers. Intralingual transfer is the strongest factor that affects non-native

learners' production of appropriate collocations. Besides, it was also found that collocations play a great role in determining the level of language proficiency among non-native speakers.

This study is underpinned by the theory of Interlanguage (IL) (Selinker 1974). The identification of the sources of errors based on IL Hypothesis validates that the linguistic structures and items in the learner language are fossilisable. This study employed Error Analysis (EA) framework proposed by Gass and Selinker (2008) to conduct the analysis. *Wordsmith Tools* software was used to generate the data for this study. This study referred to the *Oxford Collocations Dictionary* and the British National Corpus to determine if a collocation is erroneous. Nesselhauf's (2003) framework was used to classify the various types of collocational errors while Richards's (1974) and Tarone's (1981) classifications of sources of errors, namely interlingual transfer, intralingual transfer and paraphrase were used to explain the sources of collocational errors. The interlingual transfer was divided into two subcategories, namely L1 transliteration and language switch while the intralingual transfer was classified into three categories, namely false concept hypothesised, overgeneralisation and ignorance of rule restrictions. Besides, the subcomponent under the paraphrase, approximation was identified as one of the sources of collocational errors. The findings of this study indicate that of all seven types of collocational errors, the one occurring most frequently is the preposition errors. With regard to the sources of collocational errors, intralingual transfer was found to be the most prominent among the three major categories of sources of collocational errors.



## **CHAPTER 1**

### **INTRODUCTION**

#### **1.1 Introduction**

This introductory chapter discusses the background to the current study. The background introduces the topic of the study and gives information about the second language (henceforth L2) vocabulary acquisition. This chapter foregrounds the notion of vocabulary and aspects related to the acquisition of vocabulary, such as collocations and relevant statements about collocations. It also provides the research questions, research objectives, significance and scope of the present study. The theoretical framework and key definitions of important terms are also presented in this chapter.

#### **1.2 Background to the study**

Generally, vocabulary acquisition involves knowing a word in the language. Knowing a word well involves knowing several aspects about words, namely the word form, word meaning and word use (Nation, 2001). The aspect of word form refers to the language either spoken or written as well as the word parts in the language. Words that are difficult to pronounce are usually difficult to learn while words that are easy to pronounce are easier to be stored in learners' long-term memory (Ellis and Beaton, 1993; Nation, 2001). With regard to word meaning, knowing a word includes understanding the form and meaning, concept and referents as well as associations of the word. The understanding of word meaning can be achieved through the analysis of words into parts such as prefixes and suffixes which can help the learning of the words. For word use, knowing a word means knowing the grammatical functions of words and word



combinations such as collocations as well as the constraints on use such as the word frequency and appropriateness.

The central idea of knowing a word lies in the aspect of word use in which learners acquire a new word in order to use it appropriately in various contexts. In relation to the appropriate word use, it is vital to acknowledge that words are not used in isolation but are rather used as chunks such as preconstructed clauses and phrases. These chunks of language are stored in the language users' memory and they draw on the chunks when using the language. The active use of language chunks indicates that language users depend heavily on larger units of language such as collocations to express their ideas more efficiently. It shows that the main purpose of acquiring vocabulary is closely associated with the proper use of collocations. Hill (2000) states that collocation is an important vocabulary aspect that helps learners use words more fluently and proficiently. Similarly, Nation (2001) also asserts that it is the knowledge of collocations that makes native speakers sound native-like and that enables the native speakers to use the language fluently.

Evidently, the knowledge of collocations is an essential and integral part of L2 vocabulary acquisition. Reflecting on what was mentioned earlier, knowing a word means knowing how to use the appropriate grammar and collocations efficiently. Clearly, the ability to use a word appropriately is determined by one's knowledge of grammatical behaviour and collocational patterns of words (Nation, 2001). As Hill points out (2000: 52), "all the elements of natural language use are interdependent" and collocations cannot

be separated from the grammatical environment in which they occur. Ideally, the grammatical elements and lexis in a collocation should be treated as a whole given the interdependent relationship between grammar and collocations because as Lewis insightfully states (1993: vi), "language consists of grammaticalised lexis". The dichotomy between grammar and vocabulary is therefore invalid since language consists of chunks of expressions rather than individual words.

Research in L2 vocabulary acquisition is becoming more rigorous with the computerised corpus analysis. This approach has revolutionised the study of word meaning in context as well as the collocational patterns of words. Corpora consist of large collections of written and/or spoken texts are stored electronically on computers. By compiling the learner language into the various computerised learner corpora, researchers are able to investigate the learner language more precisely. This potential provides opportunity for research on L2 vocabulary knowledge as well. Rather than depending on information from case studies or self-created examples, researchers are able to use the computerised corpora of L2 learners' language to investigate the linguistic patterns and vocabulary knowledge of L2 learners.

With the availability of linguistic software in recent years, corpus-based analyses of learner language provide new insights into many areas of language structure and use (Biber et al. 1998). There is a variety of linguistic software available to generate word frequency lists, and identify specific words or word combinations in a corpus. Studies on the learner language that employ computerised corpus as well as corpus techniques



enable the researchers to discover and describe the learners' vocabulary in terms of their linguistic characteristics and patterns of authentic language use. The studies of learner corpora also disclose information about second language acquisition (henceforth SLA) mechanisms and provides a means of improving learner language (Granger, 2002). Findings derived from the analyses can help shape language pedagogy to address the needs of language learners more precisely.

With regard to vocabulary, the growing interest in learner corpora has intrigued the researchers to focus on the grammatical and lexical patterning of vocabulary, including collocations. As pointed out earlier, the knowledge of collocations has been widely recognised as an important aspect in language learning (Howarth, 1998; Hill, 2000; Nation, 2001). The appropriate use of collocations enables the learners to speak more fluently, makes their speech more comprehensible and helps them produce more native-like utterances and therefore plays a very important role in SLA (Pawley and Syder, 1983; Sinclair, 1991; Cowie, 1998; Howarth, 1998; Nation, 2001).

There have been a number of empirical studies carried out on the learners' knowledge and use of collocations. However, such popularity is still not observable in the local L2 context. This linguistic phenomenon deserves considerable attention in the local L2 context given the importance of collocational knowledge in vocabulary acquisition as well as language learning processes.

### 1.3 Statement of the problem

Although many researchers emphasise that the knowledge of collocations is of great help for the language learners to achieve fluency and proficiency, it has been recognized that the language learners often have problems with collocations owing to various reasons (Farghal and Obiedat, 1995; Gitsaki, 1997; Liu, 1999; Nesselhauf, 2003, 2005; Jukneviene, 2008). Empirical studies on the knowledge of collocations among different groups of English as a second language (henceforth ESL) or English as a foreign language (henceforth EFL) learners reveal that the learners face particular difficulty in producing the appropriate word combinations because of their lack of collocational knowledge (Bahns and Eldaw, 1993; Gitsaki, 1997; Howarth, 1998; Jukneviene, 2008). Besides, studies on the collocational error analysis indicate that collocations pose major problem for language learners as learners consistently produce various types of collocational errors (Liu, 1999; Nesselhauf, 2003, 2005; Miyakoshi, 2009). Studies on the relation between collocations and the language proficiency significantly show that there is a positive correlation between the learners' use of collocations and their written language proficiency (Zhang, 1993; Mohammad, 1998; Hsu, 2007). These studies reveal that learners who possess limited knowledge of collocations have equally lower language proficiency. Apparently, the deficiency of the knowledge of collocations is a barrier to achieve proficiency as well as fluency in the language learning. In order to overcome the problems with collocations, important aspects such as the types of errors in collocations as well as the sources of such errors should be dealt with rigorously to facilitate educators in the language teaching as well as in syllabus designs.



Although numerous studies show that collocation is central to the language learning, aspects of collocations have not been dealt with to the extent that published studies on collocations in the local L2 context are rather scarce. Therefore, the present study is an attempt to fill the research gap in collocation studies in order to better understand the collocations produced by Malaysian learners. Since there are different types of collocations in English language (refer to Section 2.6.1 for details), the present study focuses on verb-noun lexical collocations as they are the most difficult type of collocations for language learners (Bahns, 1993; Howarth, 1996; Nesselhauf, 2005; Miyakoshi, 2009).

#### **1.4 Objectives of the study**

This present study seeks to use a corpus-based method to explore a Malaysian English learner corpus by focusing on verb-noun lexical collocations. The primary aim is to identify and classify the types of collocational errors in verb-noun lexical collocations. The second aim is to investigate the possible sources of related collocational errors. The investigation of errors in collocations serves as a predictor of L2 learners' use of collocations as well as of their use of grammar as grammar and lexis (collocations) are interdependent and not separable.

#### **1.5 Research questions**

There are two research questions in the present study:

1. What types of verb-noun lexical collocational errors are found in Malaysian learners' vocabulary?

2. What sources of verb-noun lexical collocational errors are found in Malaysian learners' vocabulary?

### **1.6 Significance of the study**

By focusing on verb-noun lexical collocational errors in a computerised corpus of Malaysian L2 learner writing, the current research aims to shed some light on the collocational use among Malaysian ESL learners.

The results of the present study are anticipated to have pedagogical thrusts in facilitating the teaching and learning of vocabulary as well as the English grammar. It is hoped that by means of analysing the actual use of learner language, the study provides empirical evidence for material designers and language teachers to identify problems faced by learners in producing grammatical, proficient and natural word combinations and then offer the appropriate remedy for the problems.

### **1.7 Scope of the study**

The present study only deals with one type of word combination, which is verb-noun lexical collocation. The corpus is a subcorpus of a Malaysian learner corpus. The size of the corpus is approximately 35000 words. It only deals with the data which consists of written essays produced by Form Four Malay learners from three different states in peninsula Malaysia (Penang, Pahang and Melaka).

## 1.8 Theoretical framework of the study

This is a corpus-based study on collocations which is underpinned by the theory of Interlanguage (henceforth IL). The term *Interlanguage* was coined by Selinker (1969, 1974) to refer to the mental grammar that a learner constructs at a specific stage in the learning process. IL validates the learner language as a system of its own with its own structure (*ibid.*). IL is the product of the interaction of two linguistic systems, namely the first language (henceforth L1) and L2. IL of a L2 learner is developmental, changeable and not static. Errors are bound to occur in IL of a L2 learner since the L2 learner attempts to approach L2 by applying various rules from L2 and at the same time facing interference from his or her L1. In the present study, IL is considered as the imperfect knowledge of L2 which includes errors.

## 1.9 Definitions of terms

There are several important terms which need to be defined for clarification purposes.

### **Collocations:**

Collocations are “the occurrences of two or more words within a short space of each other in a text”. (Sinclair, 1991: 170).

Collocations are “combinations of words which occur naturally with greater than random frequency. Collocations co-occur, but not all words which co-occur are collocations” (Lewis, 1997: 25).

Collocations are word combinations which include free combinations and restricted collocations (Howarth, 1998).



Howarth's definition of collocations is adopted in the present study.

### **Corpus-based study:**

A Corpus-based study is a study which employs a collection of spoken and/or written texts stored electronically on computers as the source of data (Biber, et al., 1998; Biber and Conrad, 2001; Hunston, 2002; McCarthy, 2001; Meyer, 2002). A corpus-based study enables the researchers to compile, store as well as analyse electronically various types and different sizes of naturally occurring language data using linguistic software.

### **Error:**

An error is "a noticeable deviation from the adult grammar of a native speaker, reflecting the interlanguage competence of the learner" (Brown, 1987: 170).

An error is a reflection of a learner's "transitional competence", which is different from mistake (Corder, 1974a: 25).

### **Collocational errors:**

Collocational errors refer to the misuse of words and infringement of collocational conventions (Lombard, 1997).

Collocational errors include grammatical and lexical errors detected within the constituents of collocations (Liu, 1999).

### **Interlingual transfer:**

Interlingual transfer refers to the native language (NL) influence on the production of the learner language or Interlanguage (Corder, 1981; Brown, 1987; Gass and Selinker, 2008).

### **Intralingual transfer:**

Intralingual transfer refers to the influence within the target language (TL) learned, independent of native language (Richards, 1974; Brown, 1987; Gass and Selinker, 2008).

1.10 Summary

This chapter discussed various issues and aspects pertaining to vocabulary learning, including the importance of collocations as an integral aspect of vocabulary acquisition in SLA. The use of grammatical and appropriate collocations is the concern of the present study. Two research questions were formed in order to address the main issue raised in the present study. In the next chapter, the theoretical literature of collocations, interlanguage and errors will be discussed in detail.

## CHAPTER 2

### REVIEW OF RELATED LITERATURE

#### 2.1 Introduction

The important relationship between the knowledge on collocations and SLA can be inferred from various theoretical and empirical findings by scholars in the field of SLA. Given this, it is expedient to review the related literature before conducting the analysis pertinent to the present study. This chapter reviews the theoretical concepts and related research in collocations and collocational errors in an attempt to foreground the research gap that is addressed in the present study. The chapter culminates in a discussion of the conceptual framework of the current study.

#### 2.2 The notion of collocation

The term collocation has long been arbitrary. In the literature, scholars define collocations in various ways. The term *collocation* was coined by Firth (1957) and further developed by Halliday (1966) and Sinclair (1966, 1991). Halliday (1966: 148) states that collocation is “the co-occurrence of two words, independent of grammatical types and likely to take place over sentence boundaries” while according to Sinclair (1991: 170), collocation is “the occurrence of two or more words within a short space of each other in a text”. A short space, or “span”, is regarded as a distance of around four words (collocates) to the right and left of the node word (Sinclair, 1991: 170). If, for instance, in a corpus, the word *car* is analysed, and the words occur in an environment such as *She bought a new car from her cousin two weeks ago*, the words *she*, *bought*, *a*, *new*, *from*, *her*, *cousin* and *two* are all considered to form collocations with the node *car*; these words



are then labelled as *the collocates*. A distinction is usually made whether the co-occurrences of the words are frequent or not. Thus Sinclair's definition of collocation is based on the frequency-based approach. Some researchers (for example Halliday, 1966; Moon, 1998) adopted the frequency-based approach to define collocations as co-occurrences of all frequencies, while others (for example Kennedy, 1990; Stubbs, 1995) reserved the term for frequent co-occurrences.

Further points that are viewed differently by scholars are the number of words involved in a collocation as well as the syntactic relationship between the items in the collocations (Nesselhauf, 2005). Although Sinclair proposes "two or more words" to be considered as a collocation, the maximum length of a collocation is not set rigorously. A final aspect that is interpreted differently by researchers is the syntactic relationship between the elements. In Sinclair's frequency-based approach, it is obvious that the syntactic relationship between the elements does not determine if the co-occurrences are collocations. Co-occurrence such as *car from* is also considered as a collocation since *from* collocates with *car* within the specific span. Greenbaum (1974) excluded co-occurrences which have no syntactical and grammatical relationship such as *car from*. He emphasises that a collocation should be defined as "a single remembered set" which takes into account the syntactic and semantic relationship between the elements (Greenbaum, 1974: 80). Sinclair's definition of collocation is regarded as a looser definition compared to another approach called phraseological approach.

Based on the phraseological approach, collocation is considered as a type of word combination, which can be delimited from other types of word combinations, namely free combinations and idioms (Cowie, 1994). Cowie's definition of collocation is rather more restricted in nature. A free combination, such as *read the book*, is the least cohesive of all combinations as their components are free to combine with other lexical items. A collocation, such as *commit suicide* is more restricted in terms of its sense but less frozen than an idiom. An idiom such as *spick and span* is a truly frozen piece of language which has the least complexity. In classifying different word combinations based on the restricted sense, it should be realised that word combinations differ along a scale, which makes exact delimitation impossible. Cowie's definition of collocations based on the phraseological approach is also used variedly by researchers. Some apply the term *collocations* to all types of word combinations (for example Fan, 2009) while some reserve it for Cowie's restricted collocations while use different term such as free combinations for unrestricted combinations (for example Bahns, 1993; Farghal and Obiedat, 1995).

Another alternative definition is provided by Lewis (1997) as follows:

Besides Cowie, Howarth (1998) also proposes a definition of collocations based on phraseological approach similar to Cowie (1994). Nevertheless Howarth categorises collocations into free collocations (free combinations) and restricted collocations. He presents a collocational continuum, namely free collocations, restricted collocations, figurative idioms as well as pure idioms, as shown in Table 2.1. In this continuum, components in free collocations are substitutable. For instance, the element(s) in the combination *blow a trumpet* can be changed to *buy a trumpet*. For restricted collocations,



one of the components of the collocation has figurative meanings. For instance, the word *under* in *under attack* illustrates the figurative sense of receiving. With regard to figurative idioms, Howarth elaborates that they present the metaphorical meanings and at the same time allow the literal interpretation. Pure idioms are combinations which are frozen and cannot be predicted from the meanings of their elements. As mentioned previously, word combinations differ along a continuum, which makes exact delimitation impossible.

Table 2.1 Collocational continuum (adopted from Howarth, 1998: 28)

No.	Free collocations	Restricted collocations	Figurative idioms	Pure idioms
1	Blow a trumpet	Blow a fuse	Blow our own trumpet	Blow the gaff
2	Under the table	Under attack	Under the microscope	Under the weather

Another representative definition is provided by Lewis (1997) in his lexical approach. He defines collocations as “combinations of words which occur naturally with greater than random frequency. Collocations co-occur, but not all words which co-occur are collocations” (Lewis, 1997: 25). Lewis further denotes that collocation is arbitrary and the typical pattern of actual use of collocations can only be seen in the environments in which they have been used.



Depending on the definitions of collocation discussed above, the distinctions among collocations, idioms and free combinations are not absolute. In some cases, idioms and free combinations are included in collocations; in others, they are distinguishable. The definitions of collocations are still in a state of flux. Therefore, to provide a working definition of collocation for the present study, Howarth's definition of collocations is adopted and defined as word combinations which include free combinations and restricted collocations. This study seeks to identify the various types and sources of collocational errors in a learner corpus, thus it does not intend to investigate collocations in a semantically restricted sense. It is hoped that such a broad definition of collocation would help to gain a deeper insight into the verb-noun lexical collocational errors.

The terminology of collocation is not yet fixed. A number of labels have been given to collocations, including word combinations, lexical phrases, pre-fabricated chunks and formulaic sequences. Despite differing labels, scholars are researching on the same phenomenon, which is the role of patterning of words and word combinations in language and linguistic fields. In the present study, the term *collocations* are used more frequently.

### **2.2.1 Importance of collocations in second language acquisition**

The importance of collocations has received increasing attention in L2 learning in the past decades. Brown (1974) was among the early pioneers that emphasises the importance of collocations in language learning. He opines that the collocational

competence is vital for language learners and collocations should be emphasised in the language teaching. Besides Brown, Sinclair and Renouf (1988) view the collocational knowledge as the basis of language use and it is very important for the language learners. They should be encouraged to learn more collocations in the language classroom, especially the predictable collocations.

The role collocations play in L2 acquisition is fundamentally related to the lexical approach promoted by Lewis (1993), according to which words and word combinations are the foundations of language. The approach views the language that we use consists of many multi-word chunks, or group of words that frequently appear together. It emphasises that learners need to be exposed to word combinations in real contexts and learners should become familiar with how context affects meaning (Lewis, 1993). Lewis (1997) elaborates that fluency of a foreign language is conditioned by the acquisition of pre-fabricated chunks. Hill (2000) also stresses the role collocations play in L2 acquisition. He claims that "students do not really 'know' or 'own' a word unless they also know how that word is used, which means knowing something about its collocational field" (Hill, 2000: 60). In relation to the vocabulary learners possess, Lewis (2000) opines that the number of word combinations served as collocations is greater than the number of all words as the same words may occur in various collocations. As word combinations or chunking is very important, the lexical approach advocates the teaching of lexical phrases a primary component of the approach.



Collocational competence facilitates the learners in producing sentences or utterances which are grammatically correct as well as authentic. There is usually more than one possible way of saying something but only one or two of these ways sound natural to a native speaker of English. Often, L2 learners may produce grammatically correct utterances or sentences, yet many of them may not sound native-like, especially in producing chunked expressions. Pawley and Syder (1983) point out that native speakers of English store thousands of pre-constructed clauses in their memory and retrieve them when they use the language thus they are able to speak fluently and native-like. In relation to fluency and native-like proficiency, Lewis consistently reminds the language learners that only by expanding a range of memorised word combinations, it is possible for the learners to achieve fluency and proficiency in L2 acquisition. Similarly, Cowie (1994: 3168) states that “native-like proficiency of a language depends crucially on the knowledge of a stock of prefabricated units”. Besides, the acquisition of collocations or chunked expressions are crucial for L2 learners as it enables the learners to reduce the cognitive effort, save the processing time as well as have the language available for immediate use (Pawley and Syder, 1983; Hill, 2000; Shin and Nation, 2008). The availability of a wide range of collocations in L2 learners’ mental lexicon also facilitates as well as accelerates the communication process.

To sum up, the importance of collocations has received considerable attention in the field of SLA. The knowledge of collocations is not only the essence of a language, but also the entity that can help to increase learners’ proficiency and bring them towards native-like fluency.



## 2.3 The notion of Interlanguage (IL)

The term *Interlanguage* (henceforth IL) was coined by Selinker (1969, 1974) to refer to learners' state of the target language (henceforth TL). In the process of learning a L2 (TL), there is a continuum which stretches from the native language (henceforth NL) to the TL. The learning of TL along the continuum allows learners to adjust their NL and at the same time absorb rules and items from TL in order to approach TL. The linguistic system which expands and changes to achieve TL along the continuum is the IL system which is different from the learners' NL and also from the TL. In the present study, *TL* and *L2* as well as *NL* and *L1* are used interchangeably. No distinction is made between *TL* and *L2* as well as between *NL* and *L1*.

### 2.3.1 Interlanguage Hypothesis by Selinker (1969, 1974)

Selinker's (1969, 1974) IL Hypothesis assumes that IL is natural and systematic throughout their development and is subject to changes which occur on the continuum between NL and TL. Such a systematic development of L2 is termed *the transitional competence* by Corder (1974a). Selinker (1974) hypothesises that the linguistic items and rules in IL are likely to fossilise at certain stages of L2 acquisition. He emphasises the concept of fossilisation by stating that fossilisation is a linguistic phenomenon in which "linguistic items, rules, and subsystems which speakers of a particular NL will tend to keep in their IL relative to a particular TL, no matter what the age of the learner or amount of explanation and instruction he receives in the TL" (Selinker, 1974: 36). The fossilised structures or items are the "errors" found in IL. Selinker (1974:36) elaborates

that many fossilisable structures reappear in IL performance “even when seemingly eradicated”.

In discussing the concept of fossilisation, Selinker (1974) hypothesises five central processes which “force fossilisable material upon surface IL utterances, controlling to a very large extent the surface structures of these utterances” (Selinker, 1974: 37). The five central processes, according to Selinker, are *language transfer*, *transfer of training*, *strategies of second language learning*, *strategies of second language communication* and *overgeneralisation of TL linguistic material*.

According to Selinker (1974), *language transfer* refers to the process in which the fossilisable linguistic items or structures are resulted from NL. *Transfer of training* is the process in which the fossilisable linguistic items or structures are a result of distinguishable items in training procedures. The term *strategies of second language learning* refers to the process in which the fossilisable linguistic items or structures are attributed to the learners’ approach to the material to be learned while *strategies of second language communication* refers to the process in which fossilisable linguistic items or structures are caused by the approach learners use to communicate with native speakers of the TL. *Overgeneralisation of TL linguistic material* is the process in which the fossilisable linguistic items or structures are a result of an obvious overgeneralisation of rules and semantic properties in TL. If the errors in the learner language are attributable to the five central processes suggested by Selinker (1974), the learner language is an IL. The present study is a corpus-based study on collocations which is



underpinned by the theory of IL. Based on the findings of the current study, a conclusion will be made if Malaysian learner language exhibits the features of IL, as claimed by Selinker (1974).

### **2.3.2 Characteristics of Interlanguage by Adjemian (1976)**

Besides Selinker, Adjemian (1976) also discusses IL in terms of its characteristics. According to Adjemian, IL has three characteristics. The first one is its *systematicity*. IL is systematic and has coherent linguistic structures. Therefore, any linguistic feature of IL is analysable. The second characteristic of IL is the *permeability* of its grammar. *Permeability* refers to the susceptibility of IL to be affected by both L1 and L2 forms and rules. The third characteristic is *fossilisation*, which is non-native like competence in IL (Selinker, 1974; Adjemian, 1976; Gass and Selinker, 2008). Once the permeability of IL is lost, the IL becomes subject to fossilisation. Language learners will tend to remain certain linguistic forms or rules in their IL no matter how much pedagogical input they receive (Selinker, 1974; Adjemian, 1976; Gass and Selinker, 2008). The fossilisation of IL is the main reason to most L2 learners' failure in achieving native-like competence. In the present study, the characteristics described by Adjemian (1976) will be observed to determine if the IL of Malaysian L2 learners possesses these characteristics.

### **2.4 The notion of error**

In discussing the fossilisation of IL, it is indispensable to relate it to the notion of error as error is prevailing in the system of IL. An error refers to "a noticeable deviation from the adult grammar of a native speaker, reflecting the IL competence of the learner"



while a mistake indicates a “performance error that is either a random guess or a ‘slip’, in that it is a failure to utilise a known system correctly (Brown, 1987: 170). In SLA, an error is a reflection of a learner’s “transitional competence”, which is different from mistake (Corder, 1974a: 25). A learner’s transitional competence is best described as his current knowledge of the language, or simply his IL, based on Selinker’s definition. Error is indeed an important evidence of IL and is significant to the error analysis. The major difference between *error* and *mistake* lies in the distinction that error is systematic while the mistake is unsystematic. Errors in IL should not be viewed solely as a “product of imperfect learning” or “a reflection of faulty imitation, but rather should be viewed as indications of a learner’s attempt to figure out some system, that is to impose some regularity on language the learners are exposed to” (Gass and Selinker, 2008: 102). The present study focuses on the collocational errors in collocations. According to Lombard (1997), collocational errors refer to the misuse of words and infringement of collocational conventions. Collocational errors include grammatical and lexical errors detected within the constituent of collocations, which is synonymous to errors in collocations (Liu, 1999).

In order to study IL in terms of its errors, Error Analysis (henceforth EA) framework is employed to provide a systematic procedure for examining the IL of learners in terms of the types and sources of collocational errors. In the present study, the term *collocational errors* and *errors in collocations* are used interchangeably as both of them refer to the same systematic feature at the level of word combinations in IL. It should also be noted that the terms *inappropriate*, *deviant*, *unacceptable* and *erroneous* are used interchangeably in the present study.

## **2.5 Corpus-based study**

A corpus-based study refers to a study which employs a collection of spoken and/or written texts stored electronically on computers as the source of data (Biber, et al., 1998; Biber and Conrad, 2001; McCarthy, 2001; Hunston, 2002; Meyer, 2002). The corpus-based method of study is regarded as a refined method of finding answers to various kinds of questions researchers have always asked. Researchers are able to compile, store and analyse electronically various types and different sizes of attested language data by using the linguistic software such as *Wordsmith Tools*. Since large amounts of natural occurring language data can be accessed electronically, the study of language not only includes the structure of language but also the study of language use in real life situations (Biber et al., 1998).

### **2.5.1 Corpus-based study of learner language**

With regard to corpus-based learner language studies, Granger (2002: 4) mentions that by employing corpus-based method, “improved descriptions of learner language” can benefit various ESL and EFL studies as well as language teaching.

Corpus-based learner language analyses enable the researchers to identify areas of difficulty which learners encounter, factors that contribute to the learners’ difficulties as well as find ways to help learners develop competence in the target language (Bible et al., 1998; Nesselhauf, 2005). Nesselhauf further elaborates that “For language teaching, it is not only essential to know what native speakers typically say, but also what the typical difficulties of the learners of a certain language, or rather of certain groups of learners of



this language, are” (2004: 126). In short, corpus-based learner language studies are important for researchers and educators to understand how the language learners use the language and the difficulties they face in the acquisition of ESL or EFL.

Corpus-based studies on the ESL learner language in the local context usually focus on various grammatical categories in the language, vocabulary sizes as well as phraseological competence attained by the language learners (for example Norwati Roslim, 2004; Arshad Abd. Samad, 2006; Chau, 2008; Umi Kalthom Abd Manaf et al., 2008). Of all these research, studies on the phraseological competence such as the use of collocations by the language learners are still under-researched compared with other aspects of L2 learning, particularly in the Malaysian context.

### **2.5.2 Corpus-based study of collocations**

Sinclair (1991) claims that collocations are important as they are the basis of the language use. He also states that collocation studies significantly benefit from corpus linguistic techniques and corpus-based method of study. Corpus-based analyses enable the investigation of co-occurrence and typical context of words, such as patterns of word combinations. By using the *Concord* tool in *Wordsmith Tools*, concordance lines generated help researchers to observe the “central and typical behaviour” of a language, meaning distinctions as well as details of language use (Hunston, 2002: 42). Besides *Concord*, other tools such as *Wordlist* also contribute towards the study of collocations as it generates the word lists in alphabetical and frequency order which list out the types of vocabulary for concordance generation. *Wordlist* also generates statistical information



about the data in the studies which provides the background information about the corpus, such as the number of tokens and types of words. In short, corpus-based studies of collocations enable the researchers to access the patterns of collocations more conveniently and precisely.

## **2.6 Classification of Collocations**

Various classifications of collocations have been developed by the researchers. The present study discusses a classification system developed by Benson et al. (1986) as their classification is pertinent to the present study.

### **2.6.1 Lexical and Grammatical Collocations**

Benson et al. (1986: ix) define collocations as “fixed, identifiable, non-idiomatic phrases and constructions”. Based on the syntactic features of the words, they provide a systematic classification of collocations. They classify collocations into two major groups, namely lexical collocations and grammatical collocations. A lexical collocation typically comprises nouns, verbs, adjectives, and adverbs, for instance verb-noun combinations, adjective-noun combinations and others.

In contrast to lexical collocation, a grammatical collocation is usually a phrase made up of a dominant word (noun, adjective, or verb) and a preposition or grammatical structure like an infinitive or a clause. The present study investigates one type of lexical collocation, which is verb-noun lexical collocation. Nevertheless, following Nesselhauf's framework for classification of the types of collocational errors, the verb-noun lexical