

## Collocations in Malaysian English learners' writing: A corpus-based error analysis

ANG LENG HONG  
HAJAR ABDUL RAHIM  
TAN KIM HUA  
KHAZRIYATI SALEHUDDIN

### ABSTRACT

*Although many researchers emphasise that the knowledge of collocations is of great help for language learners to achieve fluency and proficiency, it has been widely recognized that second language learners often have problems with collocations owing to various reasons. This study describes the methods and the results of a corpus-based investigation of the types and sources of verb-noun collocational errors in a subcorpus of a Malaysian learner corpus, EMAS (The English of Malaysian School Students). The corpus consists of 130 essays written by Form Four Malay learners from three different states in peninsula Malaysia. This study was based on the Interlanguage theory and Error Analysis framework was employed to conduct the analysis. Wordsmith Tools software was used to generate the data for this study. To determine the accuracy of collocations, the Oxford Collocations Dictionary and the online British National Corpus (BNC) were referred to. Various types and sources of collocational errors were classified and explained accordingly. The findings of this study indicate that of all seven types of collocational errors, the one occurring most frequently is the preposition-related collocational 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.*

*Keywords: collocation; errors; error analysis; interlanguage; learner corpus*

### INTRODUCTION

Vocabulary acquisition involves knowing a word in the language and this involves knowing several aspects about words, namely 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 stored easily in learners' long-term memory (Nation, 2001). With regard to word meaning, knowing a word includes understanding its form and meaning, its concept and referents as well as all forms of associations with 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 word frequency and appropriateness (Nation, 2001).

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 appropriate word use, it is vital to acknowledge that words are not used in isolation but are rather used as chunks such as pre-constructed clauses and phrases. These chunks of language are stored in the language users' memory and language users 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. Other scholars (Nation, 2001; Durrant & Schmitt, 2010) also opine that it is the knowledge of collocations that makes native speakers sound native and use language fluently.

Evidently, knowledge of collocations as an essential and integral part of vocabulary acquisition contributes significantly to the Second Language Acquisition (henceforth SLA) since vocabulary learning is unquestionably central to SLA. Reflecting on what was mentioned earlier, knowing a word means knowing how to use appropriate grammar and collocations efficiently. From here, the relationship between grammar and collocations is apparent in which appropriate word use is determined by the grammatical behaviour and collocational patterns of words (Nation, 2001). The link between grammar and collocations is elaborated by Hill (2000, p.52) in which "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. The relationship between grammar and (lexis of) collocations is described insightfully by Lewis (1993, p.vi) as: "language [that] consists of grammaticalised lexis, not lexicalised grammar" in which grammar and lexis are not separable. In short, the dichotomy between grammar and vocabulary is invalid since language consists of chunks of expressions rather than individual words.

As pointed out, 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 (Sinclair, 1991; Howarth, 1998; Nation, 2001; Durrant & Schmitt, 2010). In view of the importance of collocations, a number of researchers have started to work on the definitions and boundaries of collocations. In addition, empirical studies on collocations have been carried out constantly by ESL and EFL researchers. Although studies on learners' knowledge and use of collocations are popular among ESL and EFL researchers, such popularity is still not observable in the Malaysian context. This linguistic phenomenon deserves considerable attention from Malaysian researchers since collocations are central to vocabulary acquisition and it is the most important process in learning a language (Lewis, 1993; Hill, 2000).

## LITERATURE REVIEW

Although many researchers emphasise that knowledge of collocations is of great help for language learners to achieve fluency and proficiency, it has been recognised that the language

learners often have problems with collocations owing to various reasons (Nesselhauf, 2003, 2005). Empirical studies on the knowledge of collocations among different groups of ESL or EFL learners reveal that learners face particular difficulty in producing appropriate word combinations because of their lack of collocational knowledge (Howarth, 1998). In addition, studies on collocational error analysis indicate that collocations pose major problem as learners consistently produce various types of collocational errors (Nesselhauf, 2003, 2005). 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 (Hsu, 2007). These studies reveal that learners who possess limited knowledge of collocations have equally lower language proficiency and fluency. In order to overcome the problems with collocations, important aspects such as types of errors in collocations as well as sources of such errors should be dealt with rigorously to facilitate educators in the language teaching as well as in syllabus designs.

#### APPROACHES TO COLLOCATIONS

In the literature related to this area, researchers define collocations in various ways. The term collocation was first coined by Firth (1957) and further developed by scholars such as Sinclair (1966, 1991). There have been four main approaches to collocations. The first approach is based on native speaker intuition to determine a collocation. The second approach is frequency-based in which according to Sinclair (1991), 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 relevant lexical items (collocates) of the node word (Sinclair, 1991). A distinction is usually made as to whether the co-occurrences of the words are frequent or not. Under the umbrella of frequency-based approach, the notion of collocation is further expanded. This contributes towards a more recent definition of collocation, which states collocations as the occurrence of word combinations that is greater than by chance in its context as well as word pairs that are found together more frequently than the occurrence of their component words (Stubbs, 1995).

The third approach defines collocations as combinations of particular grammatical form regardless of whether they are 'formulaic' or otherwise (Granger, 1998). The final approach to determine a collocation is phraseological-based, in which 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). Howarth (1998) also proposed a definition of collocations based on the phraseological approach similar to Cowie's, in which Howarth classified word combinations according to a collocational continuum, that is free collocations (combinations), restricted collocations, figurative idioms as well as pure idioms. 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.

The diversity of the different approaches to collocations is in fact beneficial for researchers as it provides a fruitful variety of perspectives on the phenomenon. It is thus important for researchers to consider which approach to adopt when embarking on studies on collocations. Since the study seeks to identify various types and sources of collocational errors and does not intend to examine collocations in a semantically restricted sense, Howarth's definition of collocations is adopted and defined as word combinations, which include free combinations and restricted collocations.

## THE PROBLEM

Despite occupying the status of a second language in the Malaysian education system for decades, English is still a language that has yet to be mastered by many Malaysian learners as their language is full of errors (Marlyna Maros, Tan & Salehuddin, 2007; Nor Hashimah Jalaluddin, Norsimah Mat Awal & Kesumawati Abu Bakar, 2008; Saadiyah Darus & Kaladevi, 2009; Siti Hamin Stapa & Mohd. Mustafa Izahar, 2010). Malay learners, particularly from rural schools still have difficulty using correct grammar in all aspects of language. It is worth noting that these previous studies examined various aspects of the target learner language by focusing on grammar without giving due attention to the lexical aspects of language. It would be inadequate to study the grammatical parts of a language in order to determine if one knows the language well as “all the elements of natural language use are interdependent” (Hill, 2000, p. 52). The dichotomy between grammar and lexis is invalid and therefore the analysis of both the grammar and vocabulary should be considered in the study of the language. To achieve that, it is of great importance to look at both the grammatical behaviour and the collocational patterns of words in that language. A search for the literature on collocation studies in the Malaysian L2 context indicated a near-absence with the exception of Kamariah Yunus and Su’ad Awab’s (2011) investigation of collocational competence among law undergraduates at a local university. Their research findings indicated that even law undergraduates have difficulty in dealing with preposition-related collocations. This study intends to use a corpus-based method to explore a Malaysian English learner corpus by identifying and classifying the types and sources of errors in verb-noun collocations to seek answers to the following research questions:

1. What types of verb-noun collocational errors are found in Malaysian learners’ writing?
2. What sources of verb-noun collocational errors are found in Malaysian learners’ writing?

## THEORETICAL FRAMEWORK

This is a corpus-based study on collocations, which is underpinned by the theory of Interlanguage (henceforth IL). The study will focus on the characteristics discussed in IL by Adjemian (1976). According to Adjemian, IL has three characteristics – *systematicity*, *permeability* and *fossilisation*. IL is systematic and has coherent linguistic structures. Therefore, any linguistic feature of IL is analysable. Permeability refers to the susceptibility of IL to be affected by both L1 and L2 forms and rules. Fossilisation refers to the non-native like competence in IL (Adjemian, 1976; Gass & Selinker, 2008). Once the permeability of IL is lost, the IL becomes subject to fossilisation. Language learners will tend to maintain certain linguistic forms or rules in their IL no matter how much pedagogical input they receive. The fossilisation of IL is the main reason for 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.

## METHODOLOGY

### THE CORPUS

This corpus-based study utilised a Malaysian learner corpus, EMAS (The English of Malaysian School Students), which contains both written data in the form of essays and oral interview data. The present study only investigated the written data in the corpus. The written data is an untagged learner corpus that contains data in the form of three essays written by 872 students. A subcorpus of EMAS corpus was selected for the present study, which consists of 35931 tokens of words and 2678 types of words. The selected data are a compilation of 130 picture-based essays written by Form Four Malay students. The rationale for selecting picture-based essays as the corpus data is that they are narratives and the learners would be able to express their ideas more freely thus contributing towards real language production. The average length of the essays is about 270 words.

### THE SYNTACTIC PATTERNS CONSIDERED

The present study investigated verb-noun collocations which consist of two main syntactic patterns of verbs, as shown in Table 1. In addition to examining the verb and noun combination, the grammatical elements belonging to a collocation such as prepositions and determiners (for example, articles), as suggested by Nesselhauf (2003, 2005) were also studied. The grammatical elements belonging to the verb-noun collocations were studied in order to prevent overlooking some problems of learners with regard to grammatical elements of collocations because previous researchers (Nesselhauf, 2005) suggest that collocations should be considered as a whole derived from various linguistic forms.

TABLE 1. Syntactic patterns of verbs for verb-noun collocations in the present study  
(adapted from Hunston & Francis, 2000)

| Main patterns  | Sub-patterns                    | Examples            |
|--|---------------------------------|---------------------|
| The verb is followed by a single noun group or clause  | 1) Verb noun (V n)              | Save the lady       |
|  | 2) Verb plural-noun (V pl-n)    | Saw two girls       |
|  | 3) V amount                     | Thank the three     |
| The verb is followed by a preposition and a noun group | 1) Verb prep noun<br>(V prep n) | Fall into the river |

### DATA GENERATION PROCEDURE

The method of analysis of collocational errors was based on Gass and Selinker's (2008) Error Analysis framework. Some modifications to Gass and Selinker's framework were made (as shown in Table 2) to fulfil the aim of generating data for analysis.

TABLE 2. Procedure for error analysis in the present study  
(Adapted from Gass & Selinker, 2008)

| No. | Procedure                     |
|-----|-------------------------------|
| 1   | Data generation               |
| 2   | Identification of errors      |
| 3   | Classification of errors      |
| 4   | Quantification of errors      |
| 5   | Analysis of sources of errors |

The linguistic software, *Wordsmith Tools* (version 3.0) was used to process the data. Firstly, the Wordlist tool was used to generate the statistical and frequency information about the corpus. In order to generate verb-noun collocations, the present study focuses on verbs as the node words that occur at least twice in the frequency list of the corpus. Verbs that occur only once in the corpus were omitted as proposed by Sinclair (1991). The omission caused the list of verbs to shrink to about half its size. Secondly, verbs which occur at least twice in the corpus were selected to generate their concordance lines using another tool of WordSmith Tools, the Concord. The concordance lines were generated so that all relevant verb-noun collocations could be identified.

#### IDENTIFICATION AND CLASSIFICATION OF ERRORS

After all the relevant verb-noun collocations were identified and extracted from the corpus, the Oxford Collocations Dictionary (2009) and the online British National Corpus (henceforth BNC) were used to determine the acceptability of collocations, as proposed by Nesselhauf (2005). If the word combination occurred at least five times in five different texts in the BNC, it would be accepted as an appropriate collocation. If a particular collocation was not found in the *Oxford Collocations Dictionary* and it occurred fewer than five times in the BNC, it would be considered as an erroneous collocation. After identifying the erroneous verb-noun collocations in the corpus, the errors were identified and classified into types of collocational errors based on the framework proposed by Nesselhauf (2003) as shown in Table 3.

TABLE 3. Framework for classifying types of errors in collocations in the present study  
(adapted from Nesselhauf, 2003)

| No. | Types of Errors |  |
|-----|-----------------|--|
| 1   | Verb            | Wrong choice of verb (or non-existent verb)  |
| 2   | Noun            | Wrong choice of noun (or non-existent noun)  |
| 3   | Usage 1         | Combination exists but is not used correctly                                       |
| 4   | Usage 2         | Combination does not exist and cannot be corrected by exchanging single elements   |
| 5   | Preposition     | Preposition of a prepositional verb missing, present though unacceptable, or wrong |
| 6   | Determiner      | Article missing, present though unacceptable, or wrong                             |
| 7   | Number          | Noun used in the singular instead of the plural or vice versa                      |

The identified types of verb-noun collocational errors were quantified by calculating the frequencies of the occurrences of the collocational errors and then tabulated in the forms of tables. Analysis of the data was carried out by employing a quantitative method.

#### ANALYSIS OF SOURCES OF ERRORS

The sources of verb-noun collocational errors were analysed and categorised based on the frameworks developed by Richards (1974) and Tarone (1981), as shown in Table 4. On the whole, the categorisation of sources of errors on the fundamental concepts of cognitive and communicative strategies came after the major classification of sources of collocational errors, namely interlingual transfer, intralingual transfer as well as paraphrase. The three major categorisations of sources of errors were further classified into six sub-categories of sources of collocational errors. It should be noted that the L1 transliteration and language switch were adapted from Tarone's taxonomy and reclassified into the category of interlingual transfer as the present study assigned interlingual transfer to the category of cognitive strategies.

TABLE 4. Classification of sources of collocational errors in the present study  
(adapted from Richards, 1974; Tarone, 1981)

| Strategies             | Major categorisations of sources of errors | Sub-categorisations of sources of errors  |
|------------------------|--|---|
| Cognitive Strategies   | Interlingual Transfer                      | a) L1 transliteration/L1 literal translation<br>b) Language switch  |
|                        | Intralingual Transfer                      | a) False concept hypothesised<br>b) Overgeneralisation<br>c) Ignorance of rule restrictions (failure to observe the restrictions of existing structure and analogy) |
| Communication Strategy | Paraphrase                                 | a) Approximation<br>- Semantic affinity<br>- Morphological and Phonological affinities  |

#### DATA ANALYSIS AND FINDINGS

##### TYPES OF ERRORS

Altogether, 268 erroneous verb-noun collocations were extracted from the corpus using the *Concord* tool. Some collocations contain more than one error (for example in *go for a fishing* instead of *go fishing*, a preposition error and a determiner error were counted). Thus the total number of collocational errors in the 268 collocations is 302, as shown in Table 5. As mentioned earlier, a verb-noun collocation is not only understood as consisting of a verb and a noun but also the preposition and determiner that are present in the collocation. Of all these types of collocational errors, the one occurring most frequently is the preposition errors (41.72%). This is not surprising since preposition has been found to be one of the most problematic linguistic categories in Malaysian ESL learners' writings (Saadiyah Darus & Kaladevi, 2009). The second most frequently erroneous item is the verb (16.56%), followed by noun errors (14.24%).

With regard to the preposition errors, the most noticeable ones are superfluous preposition, followed by a wrong choice of preposition and missing preposition. Superfluous preposition errors seem to be quite systematic as they mainly involve *to* and *for*. This finding indicates that the learners overuse these prepositions and at the same time they are not certain about their correct use. Besides, the superfluous preposition errors far outnumber the missing preposition errors and this phenomenon could be due to the learners' attempt to use prepositions rather than avoiding them. With regard to wrong choice of prepositions, the ones that seem to cause the most problem for learners are *in*, *to* and *into*. Learners seem to be confused by these prepositions as a great number of wrong choice of preposition errors concern the inappropriate use of these prepositions, for instance *jump in the river* and *jump to the river* (instead of *jump into the river*). This similar phenomenon was also observed by Nesselhauf (2005) in which she found that the choice between *in* and *into* pose difficulties for German learners of English.

TABLE 5. Types of collocational errors found in the learner essays

| No. | Types of collocational errors  | Examples   | Occurrences (Percentages) |
|-----|--|--|---------------------------|
| 1   | Verb   |  |                           |
|     | a) Wrong choice of verb  | a)* <i>Dropped into</i> the river<br>( <i>fell</i> into the river)             | 36 (11.92%)               |
|     | b) Non-existent verb   |  |                           |
|     | i) creation of deviant verb  | i)* <i>falled</i> in the river<br>( <i>fell</i> into the river)                | 5 (1.66%)                 |
|     | ii) misuse of item from other lexical categories                                 | ii)* <i>safe</i> that girl<br>( <i>save</i> that girl)                         | 9 (2.98%)                 |
|     |  |  | <b>50 (16.56%)</b>        |
| 2   | Noun   |  |                           |
|     | a) Wrong choice of noun  | a)* <i>come to my home</i><br>( <i>come to my house</i> )                      | 8 (2.65%)                 |
|     | b) Non-existent noun   |  |                           |
|     | i) misuse of native language   | i)* <i>has joran</i><br>( <i>has fishing rod</i> )                             | 7 (2.32%)                 |
|     | ii) misuse of item from other lexical categories                                 | ii)* <i>hear the shouted</i><br>( <i>hear the shout</i> )                      | 28 (9.27%)                |
|     |  |  | <b>43 (14.24%)</b>        |
| 3   | Usage 1  |  |                           |
|     | Combination exists but is not used correctly                                     | * <i>fall down into</i> the river<br>( <i>fall into</i> the river)             | <b>19 (6.29%)</b>         |
| 4   | Usage 2  |  |                           |
|     | Combination does not exist and cannot be corrected by exchanging single elements | * <i>story</i> about the tragedy<br>( <i>tell the story</i> about the tragedy) | <b>6 (1.99%)</b>          |
| 5   | Preposition  |  |                           |
|     | a) Preposition of a prepositional verb missing                                   | a) * <i>sat the river bank</i><br>( <i>sat on the river bank</i> )             | 13 (4.30%)                |

|   |  |   |                     |
|---|--|---|---------------------|
|   | b)Preposition of a prepositional verb unacceptable, or wrong |   |                     |
|   | i) wrong choice of preposition                               | i)* <i>fall in</i> the river<br>( <i>fall into</i> the river)   | 44 (14.57%)         |
|   | ii) superfluous preposition                                  | ii) * <i>go for</i> fishing<br>( <i>go</i> fishing)             | 69 (22.85%)         |
|   |  |   | <b>126 (41.72%)</b> |
| 6 | Determiner   |   |                     |
|   | a) Article missing   | a)* <i>went to river</i><br>( <i>went to the river</i> )        | 12 (3.97%)          |
|   | b) Article unacceptable, or wrong                            |   |                     |
|   | i) wrong choice of article                                   | i)* <i>have a courage</i><br>( <i>have the courage</i> )        | 4 (1.32%)           |
|   | ii) superfluous article                                      | ii)* <i>shouting for a help</i><br>( <i>shouting for help</i> ) | 23 (7.62%)          |
|   |  |   | <b>39 (12.91%)</b>  |
| 7 | Number   |   |                     |
|   | a) Noun used in the singular instead of the plural           | a)* <i>saw two girl</i><br>( <i>saw two girls</i> )             | 17 (5.63%)          |
|   | b) Noun used in the plural instead of the singular           | b)* <i>take this flowers</i><br>( <i>take this flower</i> )     | 2 (0.66%)           |
|   |  |   | <b>19 (6.29%)</b>   |

Total number of errors & percentages  
302 (100%)

\* = erroneous collocations

Collocations in parentheses = correct/suggested collocations

Verb errors constitute the second largest erroneous linguistic category in the current study. Of all the verb errors, the wrong choice of verbs is the most prominent verb error in the current study, followed by the misuse of item from other lexical categories as verbs and the creation of deviant verbs. It is worth noting that wrong choice of verbs is more obvious than the other two subcategories. This finding indicates that learners possess the related vocabulary (verbs) but they are not certain about the correct use of these verbs in the collocational environments. For instance, the use of *drop* and *fall* in the collocations *drop into the river* and *fall into the river* shows that *fall* is more appropriate than *drop* given the collocational restriction of the structure.

Of all noun errors, the misuse of item from other lexical categories as nouns is the most obvious noun errors followed by the wrong choice of nouns and misuse of native language. The noun errors concerning the misuse of item from other lexical categories far outnumber the other two types of noun errors, for instance *heard the shouted* (instead of *heard the shout*). Such errors indicate that the learners tend to approximate the target forms which share the similar semantic profiles with the erroneous form. This is because *shout* and *shouted* are related in terms of their meanings, although they are not synonymous with each other.

## SOURCES OF ERRORS

The sources of collocational errors were based on the errors identified in the learner language. Each collocational error was assigned to one sub-category of sources of errors. On the whole, cognitive and communication strategies are apparent as the intralingual transfer far outweighs the interlingual transfer and paraphrase strategy. Altogether, six sub-categories of sources of collocational errors were determined. They are presented in Table 6. Of all the sub-categories of sources of errors, the aspect of ignorance of rule restrictions appears to be the most influential factor which affects the learners' production of appropriate collocations in English. The results show that 59.60% of the collocational errors are attributed to the ignorance of rule restrictions, followed by approximation (21.19%), L1 transliteration (9.60%), false concept hypothesised (5.63%), and language switch (2.32%). The least influential factor is overgeneralisation, constituting only 1.66% of total errors.

TABLE 6. Sources of collocational errors in collocations

| No. | Strategies             | Sources of collocational errors                | Examples   | Occurrences (Percentages) |
|-----|------------------------|--|--|---------------------------|
| 1   | Cognitive Strategies   | Interlingual transfer                          |  |                           |
|     |                        | a) L1 transliteration / L1 literal translation | a) * <i>story</i> about the tragedy<br>( <i>tell the story</i> about the tragedy)  | 29 (9.60%)                |
|     |                        | b) Language switch                             | b) i) * <i>has joran</i><br>( <i>has fishing rod</i> )   | 7 (2.32%)                 |
|     |                        |  |  | <b>36 (11.92%)</b>        |
| 2   |                        | Intralingual transfer                          |  |                           |
|     |                        | a) False concept hypothesised                  | a) * <i>dropped</i> into the river<br>( <i>fell</i> into the river)  | 17 (5.63%)                |
|     |                        | b) Overgeneralisation                          | b) <i>heared</i> the shouted<br>( <i>heard</i> the shout)  | 5 (1.66%)                 |
|     |                        | c) Ignorance of rule restrictions              | c) <i>go for</i> fishing<br>( <i>go</i> fishing)   | 180 (59.60%)              |
|     |                        |  |  | <b>202 (66.89%)</b>       |
| 3   | Communication Strategy | Paraphrase                                     |  |                           |
|     |                        | Approximation                                  | * <i>cutting</i> some flowers<br>( <i>picking</i> some flowers)<br><br>* <i>safe</i> my friend<br>( <i>save</i> my friend) | 64 (21.19%)               |
|     |                        | Total number of errors & percentages           |  | <b>302 (100%)</b>         |

It was found that the preposition errors appear to be significantly affected by the ignorance of rule restrictions as almost all preposition errors were traced to this source. The use

of analogy as an ignorance of rule restrictions also contributes towards the inappropriate use of prepositions in which the learners attempt to use the same linguistic elements of a particular structure acquired previously on other similar structures without considering their collocational and grammatical restrictions. Similarly, the findings of Nesselhauf's (2005) study on verb-noun collocational errors also suggested that German learners of English produce preposition-related collocational errors using L2 materials in which the source of errors is intralingual.

Approximation is the second largest sub-category of the source of errors in the present study. Errors which are caused by approximation either share similar phonological and morphological features with the correct items or have semantic affinity with the target items. The results of the analysis indicate that noun errors are significantly affected by the strategy of approximation. The learners resort to the strategy of approximation as the correct structure in the target language has not been firmly formed. They also resort to the strategy of approximation when they are not certain about the correct structure. In short, learners tend to approximate to ease their linguistic burden.

In the current study, L1 transliteration took into account the word-for-word translations from L1 as well as L1 characteristics. It was found that almost all number errors are attributed to the L1 transliteration. The finding indicates that the L1 materials still play a role in the acquisition of L2, even in the L2 of the above average learners. False concept hypothesised is a result of the learners' faulty comprehension of distinctions in the TL. The learners use near-synonymous words to substitute each other without considering the collocability and contextual appropriateness of the substituting word with the neighbouring items. The confusion between the synonymous words might also contribute to the wrong choice of verb and noun. The confusion indicates that the learners could not grasp the meanings of the target word as well as the substituted word and not only the target word.

Language switch involves the direct use of the learners' native language (Malay language) without translation. Language switch strategy was found to involve only noun category. When certain terms in L2 are not available to the learners, they resort to language switch in which they transfer the whole structure of L1 into L2. Another reason for the direct use of L1 in L2 is probably resulted from the limited learners' L2 vocabulary and consequently they resort to transfer strategy from L1 to L2.

Overgeneralisation is the least common source of collocational errors found in the Malaysian learners' vocabulary and it involves the creation of deviant structures on the basis of learners' previous experience of the structures in L2. Verb errors are the only type of errors caused by overgeneralisation. These verb errors are non-existent verbs as they are ill-formed, for instance the deviant formation of past tense of irregular verb by attaching the morpheme *-ed* to the irregular verbs. Although the occurrences of these errors seem minor, such errors are too overt to be ignored and if they are fossilised in the learner language, they would not just affect the collocational competence but also the learners' basic grammatical awareness. This is because the learning of regular and irregular verbs is fundamental in ESL.

## CONCLUSION AND RECOMMENDATIONS

The present study attempted to investigate the types and sources of verb-noun collocational errors in a Malaysian learner corpus using a corpus-based method. The collocational errors

identified in the learner language were categorised according to the linguistic categories. Overall, Malaysian learner language is indeed an Interlanguage as it possesses the IL characteristics suggested by Adjemian (1976). The Malaysian learner language is systematic as there are systematic patterns of errors in it. Malaysian learner language is also proven permeable to learners' L1 and L2 as there are both interlingual and intralingual influences in the learner language. Lastly, there are tendencies for the items or structures of language to be fossilised as errors are prevalent in Malaysian learner language.

It is important for language teachers to enhance the learners' consciousness about the importance of collocations in SLA. A direct and simple way to do so is to teach them explicitly, as suggested by Hill (2000). Woolard (2000) suggests that collocations are best taught to learners when they learn new words as learners need to know how to use the new vocabulary with other words in context, not in isolations. Teachers may introduce some common collocations associated with the new vocabulary to the learners. Nesselhauf (2005) also advocates the conscious-raising activities by suggesting that learners should be made aware of the phenomenon in which word combinations differ along a scale, ranging from free combinations, collocations to idioms. In teaching collocations, language teachers may need to determine which elements of collocations deserve particular attention. Nesselhauf (2005) opines that collocations should be taught systematically and according to the degree of difficulty, from the most to the least difficult. Another feasible way to implement the teaching of collocations is by providing exercises which highlight much-neglected but common collocations such as noun+preposition+noun (Woolard, 2009). To deepen the learners' knowledge of collocations, teachers may also extend the learners' current knowledge to new knowledge (Hill, 2000). Learners should be trained to use their existing lexicon to generate more collocations as learners who are "collocationally competent ... will also be far more communicatively competent" (Hill, 2000: 62).

Learners have always been encouraged to look up the meaning of an unfamiliar word in dictionaries, but they are seldom asked to study the unfamiliar words together with other words with which they co-occur. That is why learners who usually know the meaning of a word can only use the words in a grammatically well-formed sentence and fail to produce appropriate and acceptable English. Woolard (2000) and Nesselhauf (2005) both emphasise the importance of dictionaries as a resource to gain the knowledge of collocations. Learners should be trained to use dictionaries of all kinds (traditional and collocations dictionaries) to learn how words work together. This will help the learners to express their ideas naturally and convincingly.

Lastly, language teachers are encouraged to guide the learners to use linguistic software as a tool to process the large amount of attested language data in various corpora (Lewis, 2000). In order to achieve native-like competence and fluency, learners should also be encouraged to access the native speaker corpora as a way to compare their L2 with the native speakers' L1. By using linguistic software to analyse both the native and non-native corpora, learners will be able to acquire real language and avoid producing non-native like collocations which are erroneous in various aspects (lexical, grammatical or structural).

REFERENCES

- Adjemian, C. (1976). On the nature of interlanguage systems. *Language Learning*, 26(2), 297-320.
- Cowie, A. P. (1994). Phraseology. In R. E. Asher (Ed.). *The encyclopaedia of language and linguistics* (pp. 3168-3171). Oxford: Pergamon.
- Durrant, P. & Schmitt, N. (2010). Adults learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188.
- Firth, J. R. (1957). *Papers in linguistics 1934-1951*. London: Oxford University Press.
- Gass, S. M. & Selinker, L. (2008). *Second language acquisition* (3rd edition). London: Routledge.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.). *Phraseology: Theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.
- Hill, J. (2000). Revisiting priorities: From grammatical failure to collocational success. In M. Lewis (Ed.). *Teaching collocations: Further development in the lexical approach* (pp. 47-69). London: Language Teaching.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A.P. Cowie (Ed.). *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford: Oxford University Press.
- Hsu, J. Y. (2007). Lexical collocations and their relation to the online writing of Taiwanese college English majors and non-English Majors. *Electronic Journal of Foreign Language Teaching*, 4(2), 192-209.
- Kamariah Yunus & Su'ad Awab. (2011). Collocational competence among Malaysian Law undergraduate students. *Malaysian Journal of ELT Research*, 7(1), 151-202. Retrieved on Sept 15, 2011 from <http://www.melta.org.my/Volume-7-1-2011/q?cid=93&doi=showclass>
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Hove: Language Teaching Publications.
- Lewis, M. (2000). Materials and resources for teaching collocation. In M. Lewis (Ed.). *Teaching collocations: Further development in the lexical approach* (pp. 186-204). London: Language Teaching Publications.
- Marlyna Maros, Tan Kim Hua & Salehuddin, K. (2007). Interference in learning English: Grammatical errors in English essay writing among rural Malay secondary school students in Malaysia. *e-Bangi*, 2(2), 1-15. Retrieved on Sept 16, 2011 from <http://eprints.ukm.my/114/1/Marlyna.pdf>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins Publishing Company.
- Nor Hashimah Jalaluddin, Norsimah Mat Awal & Kesumawati Abu Bakar. (2008). The mastery of English Language among lower secondary school students in Malaysia: A linguistic analysis. *European Journal of Social Sciences*, 7(2), 92-105. Retrieved on Sept 15, 2011 from [http://www.eurojournals.com/ejss\\_7\\_2\\_08.pdf](http://www.eurojournals.com/ejss_7_2_08.pdf)
- Oxford Collocations Dictionary*. (2009). Oxford: Oxford University Press.
- Richards, J. C. (1974). A non-contrastive approach to error analysis In J. C. Richards (Ed.). *Error analysis: Perspectives on second language acquisition* (pp. 172-188). London: Longman.
- Saadiah Darus & Kaladevi Subramaniam. (2009). Error analysis of the written English essays of secondary school students in Malaysia: A case study. *European Journal of Social Sciences*, 8(3), 483-495. Retrieved on 15 Sept, 2011 from [http://www.eurojournals.com/ejss\\_8\\_3\\_12.pdf](http://www.eurojournals.com/ejss_8_3_12.pdf)
- Scott, M. (1999). *WordSmith Tools*, version 3.0.
- Selinker, L. (1969). Language transfer. *General Linguistics*, 9(2), 67-90.
- Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (Eds.). *In memory of J. R. Firth* (pp. 410-430). London: Longman.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Siti Hamin Stapa & Mohd. Mustafa Izahar. (2010). Analysis of errors and subject-verb agreement among Malaysian ESL learners. *3L The Southeast Asian journal of English language studies*. Vol. 16(1), 56-73. Retrieved on Sept 16, 2011 from [http://journalarticle.ukm.my/1022/1/Siti\\_Hamin\\_Stapa.pdf](http://journalarticle.ukm.my/1022/1/Siti_Hamin_Stapa.pdf)
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. *Functions of language*. Vol. 2, 1-33.
- Sudhakaran, B. (2008). The use of prepositions among Malay adult ESL learners. Unpublished doctoral dissertation, Universiti Putra Malaysia, Serdang, Malaysia.

- Tarone, E. (1981). Some thoughts on the notion of communication strategy. *TESOL Quarterly*. Vol. 15, 285-295.
- Woolard, G. (2000). Collocation- encouraging learner independence. In M. Lewis (Ed.). *Teaching collocations: Further development in the lexical approach* (pp. 28-46). London: Language Teaching Publications.
- Woolard, G. (2009). Noticing and learning collocation. The Workshop: *LTC International College*. Vol. 1. Retrieved on Sept 16, 2011 from <http://www.scribd.com/doc/27277175/Collocations-Chunks-and-the-Corpus-A-more-Lexical-Approach>

Ang Leng Hong  
Universiti Kebangsaan Malaysia  
leng\_hong@hotmail.com