

CHAPTER ONE: INTRODUCTION

Errors are inevitable in language learning and the production of errors is a part of the learning process. Some may view errors as a stumbling block in the learning process while there are also others who think that every error should be corrected in order to prevent the learner from making the same error again. Whatever the perception may be, errors are actually a necessary part of learning a language (Norrish, 1983).

Learner errors are no longer just a necessary part in language learning. In fact, errors are now being studied by means of Error Analysis (EA) and findings from EA are used in the development and improvement of language pedagogy and teaching materials. “Perhaps, above all, it helped to make errors respectable – to force recognition that errors were not something to be avoided but were an inevitable feature of the learning process.” (Ellis, 1994: 70)

The investigation of learner language was a particularly popular pursuit, especially in the late 1960s and 1970s, which has eventually become known as ‘error analysis’ (Nesselhauf, 2005). EA research became more and more important when language researchers realized that second language learner’s errors are potentially important for the understanding of the processes of second language acquisition, and consequently the planning of courses incorporating the psychology of second language learning (Jain, 1974). Corder (1967: 25) explains that learners’ errors are significant in three different ways: 1) they tell the teacher how far the learner has progressed and what remains for him to learn, 2) to the researcher they are evidence of how a language is learned or acquired and what strategies or procedures the learner

is employing in the discovery of the language, and 3) the making of errors reveal the device the learner uses in order to learn.

Various studies which have been conducted on different aspects of learner language are proofs that findings from analysis of learner errors provide contributions to second language learning research and help teachers to understand how their learners can learn more efficiently. For instance, Richards (1971) who examined English errors produced by various second language learners suggest that the types and causes of the errors can be categorized into four aspects. The most controversial contribution is the discovery that the majority of the grammatical errors second language (L2) learners make do not reflect the learners' mother tongue (L1) but most of the errors L2 learners make indicate they are gradually building an L2 rule system (Dulay et al., 1982: 138).

After gaining much popularity and recognition, EA research suddenly found itself under attack, especially in the 1980s. There were a number of critiques – Bell (1974), Schachter and Celce-Murcia (1977), Long and Sato (1984) and Van Els et al. (1984), and most of the criticisms fall into two main categories: 1) weaknesses in methodological procedures, and 2) limitations in scope (in Ellis, 1994: 67). The weaknesses in the methodological procedures refer to the stages of analysis which lack a theoretical framework to explain the role of the errors in the process of L2 learning. The scope in EA is limited because it fails to provide a complete picture of learner language. Language researchers know what learners are incapable of. They also need to know what language learners are capable of doing too.

In the 1990s, a corpus-based methodology was introduced in the EA research to help counter the criticisms put forward. This new approach to the analysis of learner errors is called computer-aided error analysis (CEA). The CEA system is developed by Dagneaux et al. (1998) and it is based on learner corpora, which are systematic computerized collections of texts produced by language learners.

This study investigates multi-word unit (MWU) errors among Malaysian English language learners using CEA – which is based on error analysis and corpus-based approach. A current web-search and library search displays various EA studies on grammatical errors and vocabulary errors in Malaysia; but learner corpus research in CEA is limited. On top of that, there is no published work on multi-word unit (MWU) errors in learner language research, thus far, here in Malaysia. One known work on MWUs is Su'ad Awab's (1999) investigation on modal MWUs in Malaysian Memorandum of Understanding (MOUs). Therefore, using the CEA methodology, this small-scale research hopes to reveal significant findings on the MWU errors in the Malaysian English learner corpus.

1.1 Background of the Study

This is a very important section in this study as it provides background information of this study. It is crucial to understand why the English language is important in Malaysia and the current state of vocabulary teaching and learning in schools. Besides that, it is also important to understand the characteristics of Chinese-educated students as they are the subjects in this study. The last sub-section here will define multi-word units (MWUs) and explain the importance of MWUs.

1.1.1 The state and role of English in Malaysia

The English language is the second most important language in Malaysia, after Bahasa Malaysia, the official national language. English is firmly established at all levels of the Malaysian education system, a legacy of the colonial past and post-independence policies (Morais, 2001). In line with its status as a second language in the country, English is taught as a second language in all primary and secondary schools in Malaysia (Ministry of Education, 2003). The use of English is significant in certain domains such as the mass media, international business as well as in the transactions of everyday communication. With the Internet boom, English is now even more widely used as people exchange emails, chat messages and read news as well as all types of information on different websites.

The role of the English Language has been significantly enhanced in recent years with the decision taken by the Cabinet to enforce the teaching of Mathematics and Science in all primary, secondary, pre-university and university levels. The decision was taken by the government to enable students to keep abreast of rapidly expanding developments in the fields of science and technology and to help Malaysia become a centre of academic excellence.

Even though the status of the English Language is as a second language in Malaysia, it does not mean that it is a second language for the people. It may be a first language for some and it could also be a third or fourth language for others. It is common nowadays to find that many Malaysian families speak English at home and children learn English before they learn Mandarin or Malay in school. Thus, this makes English their first language regardless of their race, whether they are Malay,

Chinese or Indian. According to Zaidan Ali Jassem (1994), even though the number is not large, there are also some people in Malaysia whose first language is English and they grow up speaking it and it is the only language that they use at home. However, there are also others who learn English as a third language where the first language is Mandarin or Tamil and the second language is Bahasa Malaysia. In a multilingual society in Malaysia, it is not easy to determine the ‘mother tongue’ of a Malaysian because it may be different for every person despite the fact that he/she is a Malay, Chinese or Indian.

1.1.2 The Malaysian education system

In Malaysia, a child attends formal schooling at the age of seven for six years. After completing primary education, the child then goes on to secondary school for another five years of schooling. At the end of the fifth year, the student has to sit for a national examination – the Certificate of Education in Malaysia (or locally known as SPM – *Sijil Pelajaran Malaysia*). If the student passed the examination, s/he can go on to Form 6 (Lower) and then Form 6 (Upper), which is equivalent to pre-university level, offered at selected secondary schools. After two years, the Form 6 student will have to sit for another national examination – the Certificate of Higher Education in Malaysia (or locally known as STPM – *Sijil Tinggi Pelajaran Malaysia*), which is a pre-requisite before moving on to tertiary education.

There are basically three types of school in Malaysia: national schools, Chinese vernacular schools, and Tamil vernacular schools. The curriculum and syllabus in these three types of school are basically similar. Students also learn the same subjects except for Mandarin and Tamil. The main difference between

vernacular schools and national schools is the medium of instruction. In Chinese vernacular schools, Mandarin is the main language used for communication. Similarly, Tamil is used widely in Tamil vernacular schools. For national schools, Bahasa Malaysia is used as the medium of instruction. Therefore, a student's language exposure is very much dependent on the type of school the student goes to.

1.1.3 The Chinese educated students

For the purpose of this study, it is important to briefly describe the background of Chinese educated students as they are the main subjects in this study. For Chinese educated students, the exposure to the English Language is usually confined to the teaching and learning of the subject itself. This is because Mandarin is predominantly used in the school environment.

In Chinese vernacular schools, the teaching and learning of the English Language is only 60 minutes per week for students who are in lower primary (Standard One, Standard Two and Standard Three). For students who are in upper primary (Standard Four, Standard Five and Standard Six), the teaching and learning of the English Language is 90 minutes per week. Comparatively, in national schools, students start learning English since Standard One too, but for at least 240 minutes a week. On an average, students in national schools are exposed to the language for about 40 minutes a day, at the age of seven (Goh et al., 9 July 2009).

At the secondary school level, the learning hours, syllabus and content of the subject, are the same for both Chinese vernacular schools and national schools. The difference is that the Chinese educated students have lesser exposure to the subject during their primary school years.

1.1.4 The current state of vocabulary teaching and learning

The Curriculum Specifications for the English Language Form 5 provides a list of words as a guide for teachers in vocabulary teaching and learning. It states that “the list of words selected for teaching is drawn from a sample of the more common or high frequency words used in daily life” (Ministry of Education, 2003: 3). Some of the words given in the list are: ‘annoy’, ‘besides’, ‘confess’, ‘desire’, ‘else’, ‘fond’, ‘gradual’, ‘hunger’, ‘influence’, and ‘pride’ (refer to Appendix 9(d) for the complete word list given). In total there are 306 suggested single words. Teachers are encouraged to use these words when teaching the learning outcomes, explain the meanings in passages/context. The word list is given as a guide to the teachers so that they are aware of the range of words to be covered.

In the Malaysian vocabulary syllabus, the suggested lexical items are single words, arranged in alphabetical order, with no systematic classification of words. More importantly, what is lacking in the vocabulary syllabus is that there is no emphasis on co-occurrence of words, which is pertinent in helping learners to contextualize the meaning of words through the actual use of language. The individual word is the most basic kind of lexical item and in contrast to individual words, there are many hundreds of thousands of lexical items which are multi-word units, often carried as ‘chunks’ in our mind (Lewis, 1993). According to him, the two most important groups are collocations and institutionalized expression. These terms will be explained in section 1.1.5.

The Malaysian curriculum prescribes to Communicative Language Teaching (CLT) where the teaching and learning of the language should be meaningful and

purposeful. It is ironic that multi-word units are not emphasized in our curriculum specification. In order to help learners to learn vocabulary in a more meaningful way, word collocations should be taught so that learners are exposed to the appropriateness of word combinations. On top of that, learners would also be able to meet the challenges of acquiring a critical mass of words for use in both understanding and producing language, remembering words over time and being able to recall them readily, and developing strategies for coping with unknown words or unfamiliar uses of known words (Thornbury, 2002).

1.1.5 Multi-Word Units (MWUs)

“Ever since Firth stated that ‘You shall know a word by the company it keeps’ (Firth, 1957: 195), it has been a practice in linguistics to classify words not only on the basis of their meanings, but also on the basis of their co-occurrence with other words” (Nattinger and DeCarrico, 1992: 20). There are many names for two or more single words when they are used together. ‘Multi-word items’ or ‘multi-word units’ is the term used by Lewis (1993) to categorise two or more words which are used as a lexical item.

Under this broad category, there are three sub-categories: ‘polywords’, ‘collocations’, and ‘institutionalized expressions’. ‘Polywords’ refer to relatively short two or three words such as ‘put off’, ‘of course’ and ‘by the way’. ‘Collocations’ describe the way individual words co-occur with others. For example, the adjective ‘beautiful’ can be used to describe many nouns such as ‘girl’ and ‘dress’, but not ‘man’. ‘Institutionalized expressions’ such as ‘I was wondering if you could...’, ‘Just

a moment, please.’, and ‘Sorry to interrupt, but can I just say...’ are pragmatic in character.

Wray (2000) used the term ‘formulaic sequences’ to refer to idiomatic forms of expressions (idioms, collocations, and sentence frames) such as ‘kick the bucket’, ‘pay attention’, and ‘The point I want to stress here is...’. In Nattinger and DeCarrico (1992), ‘lexical phrases’, ‘prefabricated language’ and ‘language chunks’ are used.

Amongst all the terms used by language experts, ‘multi-word units’ is a more general term to cover all the different types of collocations. The types of collocations will be discussed further in section 2.5. The main focus of this study is the use of collocations in learner language. Collocations is one particular aspect of vocabulary learning which deserves more attention because this is one of the main difficulties students encounter in relation to new items of vocabulary as they have problems choosing the correct combination of two or more words (Bahns, 1993).

After defining multi-word units (MWUs), it is pertinent to explain the importance of learning MWUs. Nattinger and DeCarrico (1992) and Wray (2000) have a common explanation – MWUs help language learners to save effort in processing. Nattinger and DeCarrico (ibid.) argue that the use of MWUs involve less effort compared to creating a whole construction from scratch. Wray (ibid.) relates the frequency use of MWUs to the processing load – the rarer the word, the greater the processing effort.

McCarthy (1990) uses the phrase ‘mental lexicon’ to refer to the vocabulary bank in the mind. It is compared metaphorically to a dictionary or a computer in terms of the input, storage, and retrieval processes. The common idea of these processes are:

‘input’ is processing the language through reading or writing, ‘storage’ is when the language is held, kept and not lost, and ‘retrieval’ is the process of ‘calling up’ the language whenever needed.

In a dictionary or a computer, words can be arranged and stored alphabetically not only showing their meanings, but also word-class and pronunciation. The information stored is filed together. When needed, the words, together with the meanings and other necessary information, can be searched for quickly and accurately. The mental lexicon also has the input, storage and retrieval processes, but it does not function the same way as a dictionary or a computer. The human mind is vulnerable to making errors when trying to retrieve what have been stored. Therefore, if words are stored as a MWU instead of many single units, the processing effort is reduced and errors are also minimized when retrieving the language for production.

1.2 Statement of Problems

Based on the Curriculum Specifications for the English Language Form 5 (Ministry of Education, 2003), the teaching and learning of vocabulary is very much emphasized at word level only. Through the many years of teaching experience, it has been observed that learners have problems stringing the words together accurately, especially in writing. Learners have difficulty in choosing and putting the right words together to express the intended meaning in writing. Therefore, multi-word units (MWUs) is hypothesized as a problem in learners’ writing and this issue will be addressed in this study. More specifically, this study aims to investigate the MWU errors that learners make in their writing using computer-aided error analysis (CEA) methodology in a Malaysian English learner corpus.

1.3 Research Questions

It is hoped that this study will be able to produce significant findings to these research questions:

1. Which aspects of multi-word units do Malaysian English learners have problems with?
2. What patterns of MWU errors do Malaysian English learners make?
3. What are the interlingual and intralingual factors that contribute to the difficulties in the use of MWUs?

1.4 Objectives of the Study

This study on Malaysian learners of English is corpus-based and uses the CEA methodology to investigate the MWU errors in the writing of a group of Chinese educated Form 5 learners. This study aims to prove that Malaysian learners do have problems with MWUs. With the help of *WordSmith Tools (Version 3.0)*, a concordance programme, it is able to provide a more accurate frequency count of the tagged MWU errors. Through the frequency count, the MWU errors which learners have the most problem with, can also be identified in a more systematic manner.

With answers to the research questions, this study also aims to suggest changes which should be made to the curriculum specification, to cater for the vocabulary learning needs of upper secondary Form 4 and Form 5 learners. This is to prove that vocabulary learning should be beyond word level, especially at upper secondary level. In other words, collocations should be taught.

1.5 Significance of the Study

This study is significant in error analysis (EA) research as well as learner corpus research because thus far, there is no error analysis on multi-word units (MWUs) in the Malaysian learner English corpus. According to Ellis (1994), error analysis (EA) was one of the first methods used to investigate learner language and it achieved considerable popularity in the 1970s. In Malaysia, much research on the errors in learners' writing was conducted. For example, Uthia Malar Thirugrunam (2004) carried out an analysis of verb errors in written compositions of upper secondary school students and studies on lexical errors were carried out by Ong (2007), Zahira Abdullah (2003) and Woon (2003). The past EA research which had been done was mainly on grammatical errors or lexical errors. No study has been conducted yet to investigate multi-word unit (MWU) errors. Therefore, what makes this study significant is the study of MWU errors in learners' written language.

The methodology used in this study will be computer-aided. The data which have been collected as handwritten essays will be keyboarded and turned into computer-readable data. The MWU errors will then be identified and tagged. A concordance programme, *WordSmith Tools (WST)*, will be used as an aid in the analysis of the tagged errors. The details of the process will be further explained in Chapter 3.

With the concordance programme, *WST*, it aids in providing a more detailed evidence of errors in the learner corpus and thus, a more precise description of learner language. In a recent study by Gilquin et al. (2007), they find that analyzing learner corpus data is an effective way of operationalising writing difficulties. In line with

their findings, it is hoped that the outcome from this study on MWU errors can provide suggestions to educators to help learners to improve their writings. By identifying the MWU errors, teachers can determine the difficulties which are commonly faced by students and find strategies to help students deal with them. Using MWUs accurately can help students produce more polished and more native-like constructions in their writing.

1.6 Limitations of the Study

Due to time and financial constraints, as well as the requirement of this study, there are limitations in the number of subjects, the size of the corpus and the errors which are tagged for analysis. The subjects in this study are Form 5 students. They are a group of Chinese-educated students from a Chinese school in Penang. The size of the corpus is about 40,000 words, which is made up of 90 learners' writing. The errors which have been tagged are MWU errors which appear most commonly in the corpus.