A LEARNER CORPUS STUDY OF THE ARTICLE SYSTEM

SITI FATIMAH BINTI YUSOF

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

A LEARNER CORPUS STUDY OF THE ARTICLE SYSTEM

SITI FATIMAH BINTI YUSOF

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF ENGLISH AS A SECOND LANGUAGE (MESL)

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Siti Fatimah binti Yusof

Matric No: TGB 140010

Name of Degree: Masters of English as a Second Language (MESL)

Title of Dissertation : A Learner Corpus Study of the English Article System Field

of Study: Language Acquisition

I do solemnly and sincerely declare that:

    (1)   I am the sole author/writer of this Work;

    (2)   This Work is original;

    (3)   Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;

    (4)   I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

    (5)   I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;

    (6)   I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature   Date:

Subscribed and solemnly declared before,

Witness's Signature   Date:

Name:

Designation:

A LEARNER CORPUS STUDY OF THE ARTICLE SYSTEM

ABSTRACT

Studies on learner language tend to compare learners' second language with a target language. Informed by recent reconceptualization of language learning in which bilingual minds are viewed as fundamentally different from monolingual minds (see, e.g., the notion of multicompetence by Cook, 1991, 2013), the present study explores how learner language is studied as a separate and independent linguistic system. By using Antconc 3.4.4w, this study analysed a recently developed learner corpus of written data by a group of secondary school students in Malaysia, examining how language development takes place over time. Participants composed essays based on a picture prompt and changes in their language use through the study of the articles (a, an and the) were tracked over time. Findings suggest that there were changes in the way the students used language as evidenced in their texts, from the frequency of use, the distribution of patterns as well as senses.

Keywords: Learner corpus, patterns, meaning making, multicompetence

A LEARNER CORPUS STUDY OF THE ARTICLE SYSTEM

Abstrak

Kajian terhadap bahasa pelajar 'learner language' selalu membandingkan bahasa kedua dengan bahasa target pelajar. Dengan konsep baru pembelajaran bahasa dimana minda billigual adalah berbeza dari minda monolingual (lihat : konsep multikompetensi oleh Cook, 1991, 2013), kajian ini mengkaji bagaimana bahasa pelajar dikaji sebagai sistem linguistik yang berdikari dan berasingan. Dengan menggunakan Antconc 3.4.4w, kajian ini menganalisa satu korpus yang terdiri daripada karangan oleh sekumpulan pelajar sekolah menengah di Malaysia. Ia mengkaji proses perkembangan bahasa melalui sesuatu tempoh masa. Peserta menulis karangan berdasarkan gambar dan perubahan di dalam penggunaan bahasa dilihat melalui perubahan penggunaan sistem Artikel dalam bahasa Inggeris (a, an dan the). Dapatan kajian mendapati perubahan dalam cara pelajar menggunakan bahasa direkodkan dalam bidang frekuensi, corak penggunaan dan maksud.

Kata kunci : Korpus pelajar, corak bahasa, membuat maksud, multikompetensi.

DEDICATION

Dear,

Everyone who played a part in my story,

"*Thank you*"

is the least I can say to show my appreciation for everything you have done for me.

I would not be where I am today without people like you by my side.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Introduction

Natural languages such as English, Chinese and Russian are extremely complex systems (Klein & Perdue, 1997). The process of learning a second language is a lifelong process and it is definitely a hard experience for the second language learners. It is indeed a well-known fact that the process of acquiring a second language differs from the acquisition of the first language or mother tongue (Ellis, 1994). Palmer (1917) as cited in Corder, 1967) states that humans are genetically endowed with the capability to acquire and assimilate language. During childhood, there is a period whereby the human brain is the most ready to receive input and learn a particular language.

This is known as the "sensitive period" for language acquisition but it is more commonly known as the critical period (Yule, 2012). This links to the Critical Period Hypothesis coined by Lennerberg (1967). He proposes that it is possible to learn a language within a particular window of time only; in which he claims to be within the age of as early as 2 months old up to 13 years old, or in other words from birth until the child hits puberty. An internal mechanism in the human brain will enable a human infant to internalise a language structure. By learning a second language, the language learners are replacing the predisposition language structure by some other force (Corder, 1967).

However, the process to acquire the first language is a different from the second language acquisition. Although there are a number of similarities in the way learners learn and acquire the second language and acquiring their first language, the variation in situation and other factors also produce differences in the process. Most of the learners acquire their first language in a different setting with different exposure to language as compared to the second language learning experience. During the process of second

language learning or acquisition, they are also at a different level of mental and social maturity (Cook, 2003).

Corpus-based studies have been conducted longitudinally and in a cross-sectional design (see Arshad et al., 2002; Knowles and Zuraidah, 2005 and Chau, 2012). Longitudinal studies investigate learners' production of linguistic form over a period of time. Usually, longitudinal studies will involve data collection up to 2 years, while some would only involve a few months, due to time constraint and practical challenges of getting the data. This study is not an exception. I managed to gather data of 5 months to build the corpus for the purpose of this study.

On the other hand, for cross-sectional studies, the language production of the language learners is being investigated through samples of different groups of learners and later will be used to reflect individual changes. Ellis (1985) suggests that the best way to study the development of language production of the language learner is through observation over time (see Chau, 2012; Ortega, 2013) which in this case is through longitudinal studies. In the field of second language acquisition, longitudinal data have been seriously lacking in the study of second language acquisition. Yet to understand language development, examination of learner production over time is necessary.

In Malaysia, English language has been introduced decades ago during colonisation. Since attaining independence in 1957, Malaysia has been exposed to a vast change in various fields including education. Nowadays, policies related to the use of English language in the school syllabus have been implemented (see Malaysia Education Blueprint 2013-2025). The English language has long been incorporated into Malaysian school curriculum as a compulsory subject. Generally, Malaysian students are exposed to 11 years of learning English in both primary and secondary school. However, Malay language is still using as the medium of instruction.

There has been a growing body of research investigating second language acquisition among second language learners. As the second language acquisition is naturally different from the first language acquisition and there are variations between the mother tongue and the target language, studies in the field of second language acquisition have often been focusing in the area of error analysis. Traditionally, errors are regarded negatively and have to be avoided and eradicated (Cook, 2003). It was believed that the errors committed are due to the transfers that occur during the process of language learning as well as the cross linguistic features.

Corder (1967) claims that learner errors are the evidence of the process of language acquisition in the second language learners' mind. Brown and Frazer (1964) pointed out that the best way to monitor a child's progress of language learning is through systematic errors committed by them. They postulated that if the learner is using the target language in the correct manner, it could be that they are just merely repeating what they heard. Mager (1961) claims that there is a definite language system used by each language learner and there is a possibility that the system in used is different from the so-called correct system and therefore variation occurs.

These variations are regarded as errors and as the errors committed by second language learners are considered as systematic errors, hence, the variations are regarded as systematic errors. As we progress in the area of second language acquisition studies, we could see that one limitation of error analysis is we only focus on the errors committed by the second language learners and we seem to ignore the part when the second language learner gets it right. Or in a bigger picture, we actually only see half of the whole picture of the second language learners' target language performance (Tarone & Swierzbin, 2009). Other than that, the approach of error analysis in second language studies will only record the instances of which the second language learners attempt to produce difficult forms of the linguistic features. It is when they faced difficulty and eventually fail to

achieve the desired production of language. However, error analysis does not specifically identify cases whereby the second language learners avoided using the form due to the factor of difficulty. A more recent acceptance of errors in the learners' use of second language is based on the fundamental shift in perspective (see Chapter 2 for more discussion on this) from a more traditional perspective of how the second language learning takes place (Cook, 2003)

1.2 Problem Statement

To date, there are a number of studies in the language acquisition field (see Morales, 2011; Sawalmeh, 2009; Butler, 2002) comparing the learners' production of target language with the native speakers' production norm and most of the studies are conducted in cross-sectional design. In this day and age, the prominence of the learner corpus research has significantly increased. There is an abundance of studies investigating learner language by considering the use of corpus and it has left an impact on the depth on linguistic research (see Chau, 2012; Chau, 2015). Despite the growing body of research investigating second language acquisition, little attention has been given towards the area of longitudinal study using corpus. Most of the studies are concerned with the errors committed by the language learners as well as comparing the second language learners' language production with the target language norm. On the other hand, corpus-based study has been focussing on the cross-sectional design and pseudo-longitudinal design.

Hence, a comprehensive and balanced way of looking at second language learners' production is crucial. One important way to fully understand the development of learner language is not by solely looking at the errors committed by them. There is so much in the language of a language learner, rather than the errors. It is deemed as equally

important as the target language norm. The acquisition of second language learning can only be observed through thorough examination of samples of learners' performance (Ellis, 2008). This includes the analysis of samples of second language utterances or known as the learner language. Ellis (2008) explains that learner language in indeed an essential source of information in investigating second language acquisition as it provides researchers with crucial insights of second language acquisition and learner language. It would not serve justice towards the second language learners if they are just merely being viewed as deficient and incompetent.

This study is based on the view of multicompetence by Cook (2013), and translanguaging by Velasco and Garcia (2014). A bi/multilingual mind is fundamentally different from a monolingual mind (Cook, 2013). Cook stands with the notion of bi/multilingual minds and states that they must be treated as default instead of deficient language users.

Translanguaging, on the other hand looks at the dynamic process whereby bi/multilingual language users use language in a flexible manner in order to make sense of the world (Velasco & Garcia,2014). The process explains how two or more language interact and how the process can offer an impact on the process of language learning. Roughly speaking, translanguaging is concerned on the self-regulating mechanism applied by the bi/multilingual learners whereby they will be able to fully be engaged in a process of learning rather than merely participating in pedagogy (Velasco& Garcia, 2014).

Working hand in hand with the concept of multicompetence (Cook,2013) and translanguaging (Garcia, 2014) as the core idea of this study, in order to fully make sense of the language development among second language learners, a linguist must not put the production of language by language learners in comparison with any external references, but rather, the comparison would best be internal and the only yardstick is the language

itself (Selinker, 1972). In this study, I hold close to the aforementioned concept as it will only be a fair game if we view the learner language in their own form without any prejudice and preconception on how the language should be. The support to such view comes from the notion of translanguaging that believes that the language users are not using two separate linguistic systems, but rather they use both linguistic systems alternately according to the suitability of the task given to them.

The present study aims to address this concern by developing a longitudinal data set, or 'corpus', of learner language. The present study aims to add to the literature by focussing on the description of language production produced by bi/multilingual language users without any comparison to other target language norm (Cook, 1991) with the hope to add insights on the process of language development.

1.3 Objectives of the study

This study aims to examine the development of language among language learners based on the use of the English article system, as well as to add on the existing literature pertaining to language development in general and the English article system in particular. This study is based on two main objectives which are:

1. To develop a longitudinal data set or corpus of learner language.
2. To identify changes in the way learners use the English article system over time.

1.4 Research Questions

This study is guided by the following research questions:

1. What are the most frequent patterns of use of the article system at Time 1 of the learner data?

2. What are the changes in the use of the article system across the three points in time in the learner data?

3. How might such changes be explained in terms of language development?

Presented above are the research questions that I seek to find the answers throughout this study. The first question is empirically constructed as it will look at the most frequent pattern of use of the article system at the beginning of data collection time. This question will give an idea on how the students are at the beginning of data collection. Though it must be noted that this does not represent their state of language as a whole, but rather it serves as a starting point of my observation towards their language development.

The second question attempts to dig deeper by looking at the changes in the way the English article system is used throughout the three points in time. It will study how the focussed articles are identified in the corpus. This research question will look at the evidences of changes in the use of the selected English articles. The word used concurrently to the English article will be observed and the pattern of use of the English article will then be analysed internally for each time it is used in the corpus, in order to identify the frequency of use and the pattern of use of each of the article. This research question also looks at the senses of each of the article system used in the corpus. In doing so, three different dictionaries were used in order to classify the senses found in the corpus.

The third research question looks at the degree of differences to which the use of the English article system across the three points in time. Through this question, we will see an example as to how language differs over time and how much changes dictate development and take place over the stipulated time. This research question attempts to understand the natural process of human language development by looking at the changes

in the way the articles are used in the corpus as part of the second language acquisition. It sums up the three questions posed earlier by prompting to explain such changes in terms of language development.

1.5 Significance of the study

Specifically concerned with the questions raised, the impact of this study lies in its perspective that bi/multilingual language learners are different from a monolingual and thus there is a need to view them in their own right rather than as a deficient language user (Cook, 2013). It agrees with Garcia (2014) perspective on translanguaging which focuses on the dynamic process in bi/multilingual language production. In this study, the production of language by the language users were analysed as it is without any comparison to the language norm. By using the concept of multicompetence and translanguaging, this study views language learners as dynamic language users (Chau, 2012), hence, offering a new insight on the language development process.

It should be noted that this study is based on the chained ideas between multicompetence by Cook (2013) and translanguaging by Velasco & Garcia (2014) in which it focusses on the concept of treating language production of second language learners as they are. This study did not emphasis on errors. For that reason, this study aims to fill in the niche by providing insights on language learning and language development from a new perspective by utilizing the instrument of self-developed corpus which was collected over time. Instead, all instances of *a, an* and *the* evidence in the corpus, will be regarded as they are. This study will benefit teachers of ESL in Malaysia, where they will be able to have a better understanding on how students of English as second language use and develop English articles in their language learning.

1.6 Scope and Limitations of the Study

This study is intended to cover the scope within the context of the selected school. The students were selected based on convenience sampling due to the close proximity to the researcher, as the researcher is an in-service teacher teaching at the selected school. As for the analysis of the corpus, this learner corpus of narrative texts was analysed for frequency of use, patterns of use, meaning in use as evidenced from the learner corpus. Moreover, this homogenous learner corpus was collected from the writings of narrative recounts from a set of lower secondary students.

1.7 Outline of the thesis

This thesis consists of a total of five chapters. The first chapter is the Introduction (Chapter 1), followed by the Literature Review (Chapter 2), Methodology (Chapter 3). Following the three chapters is the Data Analysis (Chapter 4) and then I combined the discussion and conclusion of this study in Chapter 5, Discussion and Conclusion. The information of each chapter is explained below.

The first chapter explains about the crucial background information pertaining to this thesis. It will describe the background, the problem statement, the objectives of this study, research questions as well as the significance of this study. This chapter will also present the outline of this thesis.

The second chapter will provide reviews on the relevant literature related to this study. It will mainly be discussing about the notion of second language acquisition with the highlights on the learner language. This chapter also discusses the concept of multicompetence and translanguaging in depth before it moves on to explain about the

theoretical framework that has been the backbone of this study. This chapter outlines the niche in the research area that this study aims to fill in.

The third chapter is mainly about the research method used in this study. The necessary background information about this study will be discussed further in Chapter 3. Information about the participants, corpus construction as well as the method of analysis is reported in this chapter. Challenges in constructing the corpus will be outlined in this chapter.

Chapter 4 will present the findings of this dissertation, Data Analysis. This chapter will explain the findings recorded from the analysis of the concordance lines. The analysis of the concordance lines consists of the frequency analysis, the analysis of the pattern on use and the senses. In Chapter 5, I discuss further on the findings of the analysis. In this chapter, ideas from prominent figures in the Applied Linguistic will be out forward to support the result of the analysis. The main contribution of idea will mainly come from Cook (1991) on multicompetence and also Garcia (2004) on the notion of translanguaging. Last but not least, Chapter 5 will also serve as the concluding chapter for this thesis. It will present the pedagogical implications that this study has towards the second language acquisition as well as providing a few suggestions and recommendation for future studies to be conducted in the same area.

It is hoped that this dissertation will serve its purpose to bring in a new perspective to the study on looking at the learner language thus shedding a new light for linguists to further investigate the learner language and their development through the lens of corpus-based approach.

CHAPTER 2: LITERATURE REVIEW

This chapter starts by briefly reviewing the existing literature relevant to this study. I start by considering the concept of interlanguage before moving on to how multicompetence and the notion of translanguaging complement it. This chapter then moves on to review the learner corpus studies before it ventures to the theoretical framework that has been the backbone of this perspective and give a review on the previously conducted studies in the field.

2.1 Interlanguage

There has been a growth in the study of language over the past two decades, most of the studies are incorporated with the analysis of errors committed by the second language learners. Corder (1967) regards the occurrence of learner's errors as the evidence of the process of language acquisition in the second language learners' mind. Errors committed by second language learners are considered as systematic errors. During that era, there was an abundance of second language studies conducted on examining the errors committed by second language learners. Second language learners' production was put into comparison with the target language norm as the yardstick (see Geranpayeh, 2000; Butler, 2002).

Most of the time, the learners' production is considered as less successful attempts to reproduce the structural properties of the target language utterances (Klein & Perdue, 1997). This is due to the fact that most of SLA studies look at language as if it has a static view of language, thus, putting a fixed target of how the acquisition process should be (Klein, 1998). The second language learners' production of target language will be analysed not in relation to the target language; however, it will be compared to what is

believed to be correct and appropriate and widely accepted. Through the means of error analysis, information on the difficulties in language learning can be obtained, the language proficiency of the second language learners can be determined as well as to discover the way how people learn language as it described errors in systematic ways (Ludeling & Hirshmann, 2015).

Candling (2001) regards the Error Analysis concept as a mean to monitor and analyse the second language learners' language production and development. To date, there are still studies in second language acquisition conducted by looking at the errors committed by the second language learners. This showed the divergence towards the notion of individuality of multilingual minds. In the spectrum of multicompetence, multilingual minds are different from a monolingual, (Cook, 1992, 2003); thus, it should be treated in its own form.

Learner language in the perspective of Selinker (1972) is the production of target language by language learner. In doing so, their utterances might contain some attributes of their mother tongue, thus resulting in variation in their language production. Selinker (1972) regards their language production as a separate system. He constructs the notion of interlanguage; a transitional linguistic system which is activated by a psychological structure in the brain. Nevertheless, despite being viewed as a separate language system, Selinker (1972) did put the language produced by second language learners in comparison to either the learner's target language or their mother tongue. Hence, as the learner language is put in comparison with the target language or the mother tongue, the essence of uniqueness and individuality of the learner language might be hidden by the more dominant structure of the target language or the learners' L1 (Klein, 1998).

This was brought into discussion by Cook (1999), questioning the rational of putting the target language and the learner language in comparison to each other, when those languages are already defined as a "separate linguistic system" (Cook, 2008). This

showed the conception of setting the norm of the native speakers as the benchmark for the success in language learning and the guide to mark language competency.

Ellis (2012) views the study of interlanguage as a beginning point to explain the second language acquisition. It is viewed as the language formed by the second language learners which comprises a great number of "errors" that seem to have no association to the form of the first language or the second language (Yule, 2012). The idea that says any individual speaker of a language has a more complex "unconscious" knowledge of how language works than any other language (Yule, 2012).

Second language learners construct their own grammar system that is neither the same with their first nor second language linguistic system, but rather it owns an independent existence in the learners' mind (Cook, 2003). While Yule (2012) considered interlanguage as an interim system of the second language learners which has some features of the first and second language plus some of which are completely independent and different from the features of both languages. This shows that there is an in-between system used in the process of second language acquisition that contains traces of some aspects from the first and second language.

Languages can be separated or closely linked in the mind of a bi/multilingual; but one language may affect the other and second language learners will create their own linguistic system with unique properties (Cook, 2003). Interlanguage will naturally grow and become a space of the second language learners for communication given the appropriate and suitable conditions (Yule, 2012). A bi/multilingual mind is different form a monolingual mind. The ability of using different languages and the amount of accumulated knowledge and language input are different in monolingual and bi/multilingual minds. Grosjean (2012) states that bi/multilingual mind should not be viewed as the addition of two different monolinguals, but rather, the bi/multilingual minds possess their own linguistic features and characteristics and it is different from the art of

13

code-switching. As the learners' interlanguage develops, there will be a clear necessity for more communication and the kind of "negotiated input" that arises in conversation. Language is a dynamic and complex system, thus, the process of language learning and acquisition should be treated as a complex system on its own too.

The idea of viewing language as a dynamic system can be observed through the complexity theory coined by Larsen-Freeman (1997). It revolves around the idea that languages are able to organise itself, just like any other complex system (Larsen-Freeman, 2006). Language is also deemed as dynamic, as it is everly changing (Larsen-Freeman, 2011) and thus denying the view that language is static and there is a fixed target of how a learner language should be (Klein, 1998).

As outlined by Selinker (1967) there are five central processes in interlanguage continuum. The first one is the overgeneralisation of the target language rules whereby the second language learners overuse the rule that they learn in acquiring the second language. Some of the second language learners tend to apply the past tense morpheme – ed to all verbs to indicate the past tense form. The second central process in interlanguage is language transfer, followed by transfer-of-training, strategies of second language learning and also strategies of second language communication. Selinker (1967) also stresses the fact that underlying these five central processes of interlanguage process, there are a lot of other processes that may affect the interlanguage.

2.2 Stages in Second Language Development

Learning is generally a complicated process and it is developed step by step (Ozfidan et al.,2018). Ellis (1994) states that the process of acquiring a second language is different from the acquisition of first language or known as mother tongue. It is said that by learning a second language, language learners are replacing the predisposition of their first language by some other force (Corder, 1967).

In second language development, linguistic structures are known to develop in an established sequences regardless of the learners' native language. These sequences may overlap with each other. Corder (1967) states that in the process of second language development or acquisition, input is not equal to intake. This is because linguistic structures develop in a predictable order over time, regardless of the way learners are taught or their native language.

Second language learners are equipped with a 'built-in syllabus' whereby the learners will tend to produce the same simple linguistics structure at the beginning of the development process. Movement from one stage to another is gradual, not "all or nothing". Tarone and Sweirzbin (2009) states that there will be a transition period in between those stages. Learners are found to shift back and forth from Stage 1 to Stage 2 of acquisition as they are producing the language and the process is unique to each learner, thus creating variation.

Studies on developmental stages of language learning have been conducted both longitudinally and cross-sectionally. Longitudinal studies will examine the learners' progress over a period of time while a cross-sectional study will observe the learners' progress at a single point in time. These studies show that instructions do not seem to strongly affect the developmental sequences of learner language (see Lightbown 1983a,1983b). Instruction can speed up the sequences. However, it cannot make the

learner skip the developmental stages or alter the order of stages (see Pavesi, 1984). For instance, Pienemann (1989) in his study on the acquisition of German as a second language states that instruction did not change the natural order of which the rules are acquired. The patterns and structure produced by the instructed learners were similar to the uninstructed learners.

2.3 Multicompetence

A further advancement in thinking about learner language comes from Cook (2013) who has argued that a bi/multilingual mind is fundamentally different from a monolingual mind. He coins the term multicompetence to describe bi/multilingual minds and firmly states that they must be treated as default and not as deficient language users. It was firstly labelled as "the compound state of a mind with two grammars" (Cook, 1991). The view about learners' language production as being deficient has been challenged and any association with the native language norm has now been considered in a completely different light. There is a clear need to address the bias towards a monolingual native speaker construct in second language research and eliminating the perspective of viewing learner language as deficient (Chau, 2012).

Despite the fact that there have been studies being conducted in the field of second language acquisition with the common focus on the transfer of L1 on L2 as well as the cross linguistic features, the concept of multicompetence offers a new perspective on viewing the subject matter. In multicompetence, a bi/multilingual minds are different from a monolingual mind (Cook, 1992, 2013). The concept of treating the second language learners in their own right has changed the notion of viewing them as deficient, but rather the second language learners are now considered different from a native speaker. This situation has revealed the gap and questionable action of comparing the

language produced by the language learners and the target language norm. How can we compare the language produced by a second language learner to a monolingual native speaker?

This question provides space to reconceptualise the bias towards native speakers. The concept of multicompetence stands closely to the notion of interlanguage. It articulates the conception that if the second language users are to be treated in their right, native speakers have no particular status towards them (Cook, 2013). Hence, the essential key is the users' own language. The ability of language learners in using the language learned to convey meaning and the amount of knowledge of language accumulated by language learners have evidently showed that bi/multilingual minds are different from a monolingual native speaker. However, most of SLA studies, for instance, Geranpayeh (2000) and Butler (2002) tend not to be in coherence with multicompetence view, as those studies tend to be comparing the second language learners' language production with native speakers' norm as the yardstick.

Instead of treating the second language learners' language production as they are, they were being measured and viewed against the native speakers. The variations that exist in the second language production of second language learners' are being viewed as mistakes as it fails to conform to the language of the monolinguals. This viewpoint somehow puts the second language learners in a position whereby they will always be regarded as deficient and not competent enough as compared to the native speakers.

The concept of multicompetence has the crucial implication for language teaching. Through the lens of multicompetence, the goal of language teaching has gone through a shift whereby it is now focussed on producing a successful second language user instead of a mere imitation of native speaker (Cook, 2013). The "errors" committed by language learners are nothing but the proof of the unique separate language system embedded in bi/multilingual minds.

17

2.4 Translanguaging

Research on second language learning looks at how second language learners perform in their second language and compare their performance with native speakers' production norm. This phenomenon somehow triggers questions of the rationale of putting second language learners into comparison with native speakers of the target language. If we are comparing learner language with the native language norm, then which native language are we comparing them to (see Chau, 2012).

Velasco and Garcia (2014) claim that the understanding on how two or more language interact and how the process can offer an impact on the process of language learning has not been thoroughly enlightened and this is due to the norm that most bi/multilinguals programs tend to categorise languages in separation. This leads to the label of "two monolinguals in one". Instead, in a bigger picture, translanguaging is more concern on the self-regulating mechanism applied by the bi/multilingual learners whereby they will be able to be fully engaged in a process of learning rather than merely participating in a common pedagogy (Velasco& Garcia, 2014).

The notion of translanguaging looks at the dynamic process whereby bi/multilingual language users use language in flexible manner in order to make sense of the world (Velasco & Garcia,2014). It looks at how multiple languages relate to each other inside the mind of one person, during both process of acquiring the second language and while using it (Cook, 2003). It focusses on the flexibility and meaningful actions taken by bi/multilingual language users in order for them to communicate effectively, without viewing the languages of bi/multilinguals as separate linguistic systems (Velasco & Garcia, 2014) but instead it offers insights on how two or more languages interact and affect the learning process. From this view, the bi/multilinguals are in a situation in which

they exist in a functional relationship with the other languages that they practise and forming an integrated system of language learning.

Bi/multilinguals own only one complex system of linguistic repertoire in their mind in which they need to carefully select the appropriate features for each task. They will not have to separate their first language and the second language, but rather they own one linguistic system fully equipped with features that have been embedded in their linguistics repertoire (Velasco & Garcia, 2014). Grosjean (1989) sees a second language learners as having two modes of language use, one is the monolingual mode in which the learner actively selects the linguistic feature that they want to use, and on the other hand is the bilingual mode whereby all languages that exist in their language system is in use. It can be viewed as a strategy to win a goal, whereby in the case of translanguaging, the bi/multilinguals use the self-regulation mechanism as their strategy to achieve optimum communication by incorporating all the languages that exist in their linguistic repertoire to create meaningful communication.

It should be understood that the notion of translanguaging is beyond a mere codeswitching process as the latter is more concern in viewing and treating two languages as a separated system. Rather, translanguaging on the other side, promotes the idea that two or more languages that exist in the bi/multilingual minds forming an integrated system and are used simultaneously in order for the bi/multilinguals to achieve meaningful communication. It is not a mere communicative strategy employed by the second language learners, but rather it is made up of bilingual theory of learning.

2.5 Corpora and language studies

A corpus-based approach to linguistics relies on the database of authentic naturally occurring instances of language which is being stored in computer and being analysed

digitally through the means of corpus concordance software (Biber, Conrad & Reppen, 1998). Corpus tools have reinforced the position of descriptive linguistic as well as improving the theoretical oriented linguistic research (McEnery & Gabrielatos, 2006). There are several characteristics of a corpus-based analysis. Biber, Conrad and Rippen (1998) states that a corpus-based analysis should be empirical and using a large collection of natural texts, or corpus. It will also involve the use of computer in order to conduct the analysis. Having access to the electronic collections of L2 data can speed up the analysis and a wider range of analysis can be performed (Granger, Gilquin & Meunier, 2015). Thus, if a corpus is exploited to the fullest, it will be able to produce copious information about the language use.

A corpus is not merely a simple collection of language, either spoken or written, but rather it is a large and principle collection of natural language instances (Biber, Conrad and Rippen, 1998). It has the ability to represent a large or some part of the language with the help of a concordancing program. A learner corpus is the most ideal way to shed light on L2 word knowledge as it reveals what is missing, not yet activated or not yet produced accurately (Cobb & Horst, 2015). A corpus is planned and designed with specific purposes, which differentiate it from a common library. This specific purpose and design determines the aim and selection of texts. It is with in depth analysis by using processing data software that linguists are able to examine a language user's experience of language thus offers new perspective on the familiar (Hunston, 2010). Automated process by corpus-based software enables linguists the ability to collect, store and manage the great amount of data fast and inexpensively as well as giving the opportunity to replicate the corpus studies (McEnery & Gabrielatos, 2006). As Sinclair (1991) states that, every detail of a text has to be thoroughly examined. One of the ways to analyse learner language is through corpus driven study as corpus research has been proven the most productive in

examining and revealing the characteristics of learner language, both qualitatively and quantitatively (Chau, 2012).

Concordance software is a program that allows user to discern a specific target word in a corpus and it has the ability to carry out collection of analysis on the corpus (Biber, Conrad and Rippen, 1998). With the use of corpus concordance software, it is made possible for linguists to carry out research on the use of language especially the second language production of second language learners by enabling the data to be quantified and compared in systematic ways (Adel, 2015). Moreover, it has opened up the space for expansion of the scope of study pertaining to language use.

The optimized use of corpus-based study will add values on the study of individual linguistic features as well as the characterization of language varieties among language learners. It will be able to provide a fresh perspective on the notion of language use as it will provide systematic patterns on how the language is used. With the tools engraved in the corpus concordance software, the corpus-based analysis will enlighten the field of second language acquisition on the identification of the crucial characteristic in language use. The ability to study large text corpora in an organized manner permits access to a quality of evidence that has not been accessible before (Sinclair, 1991).

Learner language is not only known through its errors. The language has many other interesting structures that the computer can help to reveal (1994). Granger (2003) states when corpus linguistics made its appearance in 1950s, it was a very modest enterprise catered by small group of enthusiasts. Today, the discipline has expanded considerably and various studies have been conducted through the lens of corpus linguistics. Sardinha (2013) states that corpus linguistic is concerned with many aspects and one of them is lexicogrammar. Lexicogrammar is a level of linguistic structure whereby lexis or vocabulary, and grammar or syntax, are combined into one. At this level,

grammatical structure and words are being regarded as mutually dependent with one level interfacing another.

Hunston (2002) firmly states that a corpus on its own is nothing than a place to store the evidence of language because it can do nothing at all. However, the corpus access software can rearrange the language stored and through the rearrangement of the language, various observations of the language will emerge. She states that it is equally essential to note that a corpus on its own cannot and will not contain new information about the language; instead, corpus concordance software may help to provide a fresh approach towards the language. The corpus concordance software may process data in three ways through presentation of frequency, phraseology as well as collocation (Hunston, 2002).

It has been further explained that the words in a corpus can be regarded in order of their frequency in the particular corpus. The frequency list will be valuable in recognizing the possible distinctions among other corpora. Secondly, if a corpus is studied based on phraseology, the concordance lines will bring together many examples of the use of a word or phrase, constructing possibilities and space for linguists to analyse the language thoroughly. Finally, yet importantly, the collocation in which it looks at the tendency of words to co-occur together. Nowadays, corpora have a wide variety of use. When it comes to language teaching, the learner corpora will provide evidence on the mechanism of a language and how it actually works.

On this day, there is an abundance of availability of the concordance software and the number of corpus-based studies has been increasing gradually. The corpora and the corpus-based studies have indeed increased the interest towards studies in the language use, especially the second language learner's production of language. The findings of any corpus based study depends on what the researcher is looking for and the way it is analysed.

An example of corpus study is the one conducted by Bestgen and Granger (2014). This study looks at the development of phraseological competence in L2 English writing. The aim of this study is to investigate whether or not phraseological competence in L2 develops over time. The study puts the main focus on the aspect of L2 writing development among language learners. Utilizing 171 essays written by 57 learners of English from Michigan State University Corpus of ESL Writing, this study incorporates both longitudinal and cross-sectional approach. As the result, it was evident that phraseology does play a role in the development of L2 writing.

Some examples of leaner corpora in Malaysia are the EMAS corpus (English of Malaysian School Student) corpus (Arshad et al.,2002), MACLE (Malaysian Corpus of Learner English) (Knowles and Zuraidah, 2005) and CALES (Corpus Archive of Learner English Sabah-Sarawak) (Botley et al., 2005) and another corpus network, CoNLoPSEM (The Corpus Network of Longitudinal Projects on Student English in Malaysia).

An excellent example of language study involving a corpus-based approach is by Arshad (2004) whereby he investigates language development based on the EMAS corpus. The study uses the production of language and the variation in lexical items as the evidence of development in the selected corpus. This is a cross-sectional study whereby the data in the corpus was collected from three different age groups ranging from 11 years old, 13 years old and 16 years old. The learners were required to write essays based on picture series and titles that were assigned to them in three sessions of data collection. Findings from this study reveal the developmental pattern as well as the vocabulary level of the students through comparison made between the three age groups. It was evidenced through the concordance lines that there are indeed some developments in the language production of the leaners as well as their vocabulary level when compared across the three age groups.

Another study conducted in Malaysian context is by Sarimah and Nurul Ros Adira (2010) whereby they conducted an error analysis by using the corpus-based approach. The selected corpus was developed from samples of written paragraphs collected from 66 first year students of UTM. On the other hand, this study also incorporated a survey method that aimed to investigate the type of errors committed by the students in their writing and the survey was intended to enquire about how much do the students acknowledged about the errors that they commit in their writing. Findings from this study concluded that the first year UTM students still produce grammatical and lexical errors in their writing. Other than that, the analysis of the questionnaires concluded that despite knowing the types of errors committed, there are still recurrences of the same type of errors in their writing. From this study, it was discerned that the most common error committed by the students in their writing is the error pertaining to tenses.

Based on the aforementioned studies, it could be understood that there are several divisions of the corpus-based study. Language studies involving the use of learner corpora could be approached either through cross-sectional or longitudinal design. The scope to such study is also wide as language development is best observed through the lens of corpus-based studies as they provide the information on how a specific word or a grammatical item is used (Arshad, 2002). Research in English Language has received ample benefits from the corpus-based approach (Hajar, 2014), however, many studies concerning the learner corpora tend to look at the learners' error when it is very crucial to focus on treating the learners' second language production in their own right rather than putting them into comparison with the native speakers' production norm (Selinker, 1972).

2.6 English article system

In the study of language, some of the most interesting observations are made not in terms of the component of the language, but it was found in terms of the way how the language is used (Yule, 2012). In the perspective of traditional analysis of language use, the analysis of language must adhere to the "all and only" criterion of the correct language use (Yule, 2012). This means that we have to, by all means, get the correct grammatical phrases and no ill-formed structures are accepted. The term "article", "adjective" and "noun" are the terms that we use to label the grammatical categories of the word. Basic definition of this type is crucial for identifying the forms in a language such as English, Arabic, Latin and many more. However, an alternative way of examining the parts of speech used is by thorough examination on how the words are used.

English articles are among the most frequently occurring free morpheme in English (Master, 1997). It is divided into two categories, which are the definite, and indefinite articles. The definite articles consist of the English article *the* while the indefinite articles consist of the English articles *a* and *an*. The COBUILD (Collins Birmingham University International Language Database) indicates that the article *'the'* is the most frequently identified word in the corpus while the article *'a'* holds the fifth position after *'of'*, '*and'* and '*to'*. The English articles, *a, an* and *the* in Yule (2012) point of view are the words which are used with nouns in order to form noun phrases in which it has the ability to classify those "things", for instance, *You can have a banana or an apple.* In the aspect of English articles, there are three main elements that comprised it which are countability, number and definiteness (Master, 1997). All these three elements must be considered thoroughly in examining the learners' use of the selected English articles.

On the other hand, the English articles are also used in identifying the noun as they are already known, for instance, *I'll take the book.* Scholars had discussed the acquisition of the English articles whereby Master (1997) posits that the acquisition of the English articles *a* occur autonomously of *the*. The acquisition of the English articles *a* was distinctive due to the fact that it was not found to be influenced by instruction. Master (1997) also states that in general, each of the English articles has their own functions. The indefinite English articles *a*, generally function to signify a boundary that makes a formless entity discrete and thereby countable. The English articles *a* tend to co-occur with singular count nouns and is the second most common way to describe a generic noun (Master, 1997).

On the other hand, the English article *the* is used to single out, identify and indicate that the speaker either wants to single out a noun and identify it for the hearer or to instruct the hearer to do so (Master, 1997). Most of previous studies on English article system are commonly focussed on the errors committed by the second language learners. One instance of such study was conducted by Morales (2011) whereby he compared the acquisition of English article system by L1 Spanish speakers with native speakers. Findings from the study shows that native speakers performed better as compared to second language learners who were able to incorporate 92% of correct usage of article while mixing it up with some omission and errors. Another study by Miller (2005) has articulated the importance of countability and the ability to address specific use of article among Chinese L1 speakers studying in an Australian university.

The participants of this study were 41 university students studying in an Australian university from different races and background. Two tests were administered to the participants. The first one was the gap-filling test and the second one required the participants to correctly identify the missing articles in a short paragraph. Afterwards, they were exposed to a lesson of English article system and The single-session

experimental study has proven that with reinforcement of the notion of countability has led to improvement in accuracy as it helps to increase leaners' awareness in identifying nouns that are countable and uncountable.

Sawalmeh (2009) looks at the aspect of errors committed by second language learners in their writing. Following Corder's (1967) error analysis framework, essays written by Arabic speaking learners of English were analysed and it was observed that the learners committed errors in 10 aspects of English language and one of them is the English article system which takes up to 12.4% of the errors. In a similar vein, Butler (2002) looks into the notion of English article system development of eighty Japanese college students at different stage of interlanguage development. The participants were required to answer a set of filling the gap test whereby they were required to fill in the appropriate articles for each gap. Afterwards, they were required to provide justification for their choice of articles in the gap-filling task. The findings from the study concluded that learners' errors arise from a number of causes and learners have different styles to understand English articles. Both of the aforementioned studies show the importance of achieving the native speakers' standard when using English article system.

Another study that also stands on the importance of achieving correct standard of English article system is conducted by Geranpayeh (2000) in which he investigates the challenges faced by Persian speakers in acquiring the English article system. 15 postgraduate students studying in Edinburgh and Newcastle participated in the study. Similar to Butler (2002), data required for the study was gathered through two tasks of which the participants were required to complete a gap-filling test by filling in the missing articles. On the other hand, the second task tested the participants in their error correction ability whereby they were tested on their sensitivity towards the English article system. For comparison purposes, the same tests were distributed to a group of native speakers and the result obtained were put into comparison. Results revealed that Persian learners

had difficulties in identifying the English definite marker when it was placed in the subject position and it was discerned that the L1 transfer was evidenced when the articles appear in the subject position.

I attempt to break through this norm by concentrating on the notion of viewing learners' production without comparing them with native speakers' norm as the yardstick. As being put forward by Cook (2013) a bi/multilingual mind is fundamentally different from a monolingual mind. This study views learners' production of English articles; *a, an,* and *the* with a focus on treating learners' production in their own right without any comparison made prior to the nature of the target language norm in coherence with the notion of multicompetence by Cook (2013).

2.7 Theoretical Framework
Idiom Principle and Open Choice Principle

Language is a complex system and the meaning of any word is derived from it's discourse and not only from where it came from (Sinclair, 2004). Any instances of language will depend on its surrounding context (Sinclair,1991). In order to interpret the meaning underlying any text, Sinclair (1991) suggests two different principles of interpretation. The two principles that have been proposed by him are the open choice principle and the idiom principle. Language users will alternate between incorporating the open choice principle and the idiom principle in their language production. He states that any group of sequence of sentences will be constructed and interpreted through the means of either the idiom principle or the open choice principle, but not both at the same time. The two models of interpretation must be employed in separation as both principles are incompatible with each other. In another word, it could be discerned that the meanings of

phrases can either be discerned through the whole phrase or by the individual words contained in the phrase.

The idiom principle describes that things that tend to physically occur together have a stronger chance of being said or to occur together (Sinclair, 1991). Words are not randomly occurring in a text and it is not possible to produce a normal text by only depending on the open choice principle. In a similar vein, Hunston (2002) posits the same idea by defining the idiom principle as seeing each and every word in a text as having its meaning attached to the whole phrase rather than to the individual parts of it. And due to this, the hearer or the reader will be able to discern the meaning of a phrase as a phrase and not as a chunk of a grammatical lexis. It can be understood that in the light of the idiom principle, language users have a great amount of semi-preconstructed phrases that reflect single choices, even though as a glance the semi-preconstructed phrases might seem to be interpreted and parsed into segments (Sinclair, 1991).

Hunston (2002) also claims that if a text cannot be interpreted and understood through the concept of the idiom principle, the language user will bounce back to the concept of the open choice principle. In contrast to the idiom principle, the open choice principle views language as having a very large set of complex choices with a large range of choices opened up for use and the only limit is the grammaticalness (Sinclair, 1991). This principle is also known as the 'slot-and-filler' model illustrating the texts as a series of empty slots which have to be filled from a lexicon that complements the restraints. Any word may fill in the empty slots.

This data driven study views the second language learners' production of language as it is without imposing any own ideas on it (see Chau, 2015) bringing up the notion of trusting the text (Sinclair, 2004) complemented with multicompetence view by Cook (2013). The concept of treating the second language learners' production of target language in their own right is utilised in this study.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter will present and discuss the research methods involved in this study. It will start by explaining the location of which the learner corpus was gathered with the information of the selected school. Then, this chapter will explain the participants who have contributed in this study through their written texts as well as the procedures and instrument involved during the data collection process. A brief review on the gathered learner corpus, the CoMENT, is also provided in this chapter. Towards the end of this chapter, the analytic procedure employed in this study is explained.

This study analyses a corpus of learner English narrative texts written by secondary school students in Malaysia which was collected over time. It employs a time-series design with features of a longitudinal study. The corpus was analysed using the concordance software Antconc 3.4.4w (Windows) 2014. To date, the corpus-based approach is widely used to study linguistic patterning to the learner language and analysis is used to study the changes in the structure in second language acquisition studies.

3.2 Location

The location of the study was a secondary school in Temerloh. The school runs in a single session with approximately 311 students. The Malay language is used as the medium of instruction whereas the English language is taught as a compulsory subject within the school curriculum. All of the participants experienced approximately the same number of eleven years of education, through primary and secondary education system.

The distribution of the students of the school by form and ethnic background is shown in Table 3.1 below.

Table 3. 1 The Distribution of Students by Forms and Ethnic Backgrounds

Form (Gender)	Malays		Orang Asli		Total	
	M	F	M	F	M	F
1	23	31	7	7	30	38
2	20	29	4	8	24	37
3	31	36	5	7	38	41
4	20	23	3	1	23	24
5	23	26	4	3	27	29
Total	117	145	23	26	140	171

(Source: School Registration Record of November 2016)

3.3 Participants

A total of 35 students contributed their narrative texts to the development of the longitudinal corpus used in this thesis. The initial number of participants was 41 students; however, the project experienced attrition in the number of participants over time. The students were Secondary One (13-year-old) students when the data collection began in November 2015. Among the 35 students, 21 are female and 14 are male. They speak the Malay language as their mother tongue at home and they learn English at school.

As for the Orang Asli students, they came from Jah Hut tribe, one of the Orang Asli tribes living in Pahang. This tribe lives around a hill tract area, ten to twenty miles from the river bank of Pahang river, between Jerantut and Temerloh (Diffloth,1976). They speak in their own ethnic language known as Jah Hut language at home with their family and relatives and they learn and use Malay and English at school.

3.4 Procedure

During each data collection session, all of the participants completed a writing assignment that involved essay writing. They were required to write a narrative essay based on the picture prompt of a group of boys saving a drowning girl by a river or lake. The time given was one hour and they were instructed to produce a minimum of 100 words narrative recount. This task follows the English format of the PT3 examination which every student in lower secondary will have to take when they are in Form 3. The task given to them was a guided writing and there were 10 helping verbs given to the participants to add in the process of the essay writing. (see Appendix A for the picture prompt).

3.5 Instrument

The main source of data used to seek answers to the postulated research questions is the written narrative essays of 35 participants of the selected school. The concordance software, Antconc, was utilised in analysing the written essays. Antconc is a program developed by Laurence Anthony. It was a very useful tool in investigating the concordance lines. After the participants' essays were typed in electronic forms, the essays were loaded into the concordance software to generate concordance lines to be used for further analysis.

3.6 The CoMENT

The **C**orpus **o**f **M**alaysian **E**nglish **N**arrative **T**exts (henceforth, the CoMENT) is a homogenous collection of texts consisting of narrative texts contributed by 35 students from a secondary school situated in Pahang. The data was obtained through three sessions of data collection of a six-months span which commenced in November 2015, early February 2016 and May 2016.

The CoMENT consists of 105 narrative texts with a total of 12,245 words. Each text is coded according to the student who wrote the text and when it was written. The coding range is 101-141. This coding system also includes the letters 'a', 'b' and 'c' to indicate the three different points in the time of the narrative texts were collected. Therefore, 101a and 102a, for example, refer to texts written by two different students ('101' and '102') and were written and collected at the same time ('a'- which is November 2015). On the other hand, 001a and 001c refer to two different narrative texts written by the same student ('001') and were written and collected at two different points in time, with 'a' referring to November 2015 and 'c' referring to May 2016. Participants were allocated one hour to write the essay during each data collection session.

During each data collection session, the participants were provided with a short list of known words that they may use in their narrative text. Data collection was done in the class during the English language lesson. During each data collection session, the researcher was present as she was their subject teacher.

The task requires the participants to write about a female character who had accidentally fallen into a river and the participants were required to describe how the character was saved from drowning (see Appendix A). Samples of the narrative text from the corpus written by participant 101a and 112a are shown below:

One day, Siti and Hafizah walked to the river for plucked flower. Aidil, Hafiz and Zaid going to fish at the river. Have a shady tree at the river. Suddenly, Hafizah fall into river. Siti shouted panicked. After that, Siti called for help. Aidil, Hafiz and Zaid shocked. Aidil, Hafiz and Zaid heard the scream. Aidil rescued the girl. Aidil jumped into the river to save the girl. Aidil, Hafiz and Zaid brought the girl to the hospital. Doktor checked and gave medicines. Hafizah said 'thank you' to the Aidil for helped her. Siti and Hafizah was happy as a lark.

(101a)

One day, Ali, Abu and Atan go to fishing. The boys walking to the river and carrying a fishing rods and pain. Amira and Arina are plucking flowers near the river. Suddenly, Arina was fall and drowned into the river. Amira was shouted for help because Arina fall into the river. Ali, Abu and Atan were shocked because heard the shouted. After that, Ali, Abu and Atan ran to the river and Atan jumped into the river for saved the girl. After saved the girl, Abu called the ambulance. The ambulance came and took the girl to the hospital. When arrived to the hospital, doctor checked the girl and gave medicine. After that, Arina said thank you to the boys because saving her.

(112a)

The same task was given to the participants from Time 1 to Time 3 of data collection. The narrative texts collected were typed into Microsoft Words by the researcher and was proof read by an independent reader in order to account for typing errors and to ensure that everything that is typed is similar to the original narrative texts written by the students. The file was then saved in .txt format so that it could be run through the concordancing software. It must be noted that all the 105 narrative texts in this corpus are all authentic texts with the conventionally labelled 'spelling errors' are

34

kept as they are. The data were collected and used for research purposes with the consent of both students and their respective parents or guardian (see Appendix for ethical consent). The participants were informed about the nature of the study by the researcher and they were free to withdraw from the study at their own will. They were ensured that their identity will be kept anonymous during and after the completion of the study. The details of the corpus are presented in Table 3.2 below.

Table 3. 2 The CoMENT

Time	No of students	No of texts	No of word tokens	No of word types
1 (Nov 2015)	35	35	4445	436
2 (Feb 2016)	35	35	3910	391
3 (May 2016)	35	35	3890	465
Total	**35**	**105**	**12245**	**1292**

3.7 Analytic procedures

This data-driven study uses the corpus analysis toolkit software, AntConc for Windows 3.4.3w (Windows) 2014 in analysing the collected data. This study focusses on single word as a starting point to understand the learner language. Chapter 4 of this thesis will present the analysis of the English articles ie: articles *a, an* and *the* from the corpus. As will be presented in Chapter 4, the analysis of the English article *a, an* and *the* can be observed through the frequency list generated as well as the pattern of use over time. The patterns of use were observed through concordance lines that were generated by the concordance software.   On the other hand, in analysing the development of the learner language, I will also discuss the variation of ways of which

learners incorporated the use of English articles *a, an,* and *the* in their writing in terms of senses.

The change in the learners' production of the selected English articles was monitored prior to answering the research questions. It should be noted that this study follows the notion of multicompetence (see Cook, 1991,2013) whereby it highlights second language learners as having their own separate linguistic system; therefore any occurrences of errors found in the learners' data will be disregarded (see Chau, 2012, 2015). Instead, all occurrences of the English articles *a, an* and *the* identified in the learners' data were taken into analysis.

As the classification of pattern can be complicated, it was done following two criteria drawn in determining whether any occurrence of the English articles *a, an* and *the* can be considered as a pattern or otherwise. First, in order to be considered as a pattern, any sequence of pattern must occur more than once in the whole corpus. Secondly, it must be used by at least two different participants throughout the learner corpus (Chau, 2015). If the sequence of pattern does not fulfil any of these two criterions, it will not be considered as a pattern and will be disregarded. Based on the analysis of data, it was revealed that there was no occurrence of the English article *an* that fits into the aforementioned criteria.

In order to determine the senses of the English articles *a, an* and *the* in the COMENT, two dictionaries have been used. The first dictionary is the Collin Cobuild Advanced Learner's Dictionary, 8th Edition, published by HarperCollins Publishers in 2014. The second dictionary is the Oxford Advanced Learner's Dictionary, International Student's Edition, 9th Edition, 2015, published by the Oxford University Press. For each pattern, the identification of the senses of the English articles *a,* and *the* was done manually through three times of analysis. Table 3.3 below shows the frequency of use of the English articles *a* and *the* observed in the COMENT.

Table 3. 3 The Frequency of Use of the English articles *a* and *the* across the Corpus

Words	Frequency			Total
	TIME 1	TIME 2	TIME 3	
A	18	42	47	107
The	509	409	340	1258

3.8 Conclusion

The analysis to be presented and discussed in the next two chapters is drawn upon the narrative texts written by 35 students participating in this study. The analysis is concerned with tracking the language development among second language learners with focus a on the English articles. The analysis was completed manually with the aid of the corpus concordance software, AntConc.

CHAPTER 4: DATA ANALYSIS

4.1 Introduction

This chapter will present and discuss the findings discerned from the analysis of the concordance lines of the learner corpus, the CoMENT. The analysis was done on three aspects which are;

i-   the frequency of use of the English articles

ii-   the pattern of use which focuses on the different variations of patterns of how the second language learners incorporate the use of English articles in their writing

iii- the senses of the English articles in which the way second language learners use English articles in their writing and the variations of inferred meaning from their writing.

All the analysis was done based on the concordance lines. The outcome of the analysis of the concordance lines results in 107 instances of the English article *a* and 1528 instances of the English article *the* in the longitudinal data, and it is discussed in this chapter as an attempt to gain further insights into the process of second language development. As noted in Chapter 3, the English article *an* was not observed in the three sets of data collection. Since there was no instances of occurrences of the English article *an* across the learner corpus. This chapter will mainly focus on the analysis of the English articles *a* and *the* only.

4.2 Analysis of the English article *a*: Frequency of use across periods of language development

In order to examine the language development of the second language learners in the learner corpus, one of the aspects taken into consideration was the frequency of use of the selected English article. The CoMENT was run through the concordance software in order to generate the frequency list as shown in Table 4.1. As can be discerned from the table below, it can be observed that the frequency of use for the English article *a* has undergone an increment over time throughout the data collection sessions.

Table 4. 1 Distribution of Frequency of Use of English article *a* in the Corpus

English article	TIME 1		TIME 2		TIME 3	
	Frequency	%	Frequency	%	Frequency	%
A	18	0.4	42	1.1	47	1.21

The frequency list in Table 4.1 suggests an increase in the rank order of the English article *a*, and it is reflected across three sets of data. As can be discerned from the table, it could be seen that the frequency of the English articles *a* was initially recorded to be used 18 times at 0.4% in the learner corpus at time 1. However, over time the frequency of use of the English articles *a* was observed to increase to 42 times in time 2 at 1.1% and 47 times, 1.21% during the third data collection session. It can be said that the frequency of use of the English article *a* in the CoMENT increased over time as can be seen by the increase in the percentage of use of the English article *a* in the corpus. The percentage of use of English article *a* was recorded the highest at Time 3 at 1.21%. This figure suggests that language learners produce more instances of indefinite articles *a* over time.

Table 4.1.1 below shows the comparison of word frequency of the top 20 words in the corpus. As can be seen in Table 4.1.1, the English article *a* was not in the top 20 most frequent words in observed at Time 1, instead it was recorded as the 54th word in the frequency list. However, beginning Time 2, the English article *a* was observed to place the 14th most frequent word and during the third data collection, it was observed to be the 11th most frequent word. It can be said that the frequency of use of the English article *a* in the CoMENT increased over time as can be seen from the rank order in Table 4.1.1. The English article *a* was not listed in Top 20 of the most frequent word in the CoMENT at Time 1. However, beginning Time 2, English article *a* was recorded at the 14th out of 20 top words in the corpus followed by Time 3 at 11th position. This situation suggests that language learners produce more instances of indefinite articles *a* over time. It is evident that the frequency of use of the indefinite article *a* has increased over time.

Table 4.1. 1 Comparison of word frequency of the top 20 words in the corpus

Rank Order	Time 1 November 2015		Time 2 February 2016		Time 3 May 2016	
1	509	the	409	the	340	the
2	270	and	231	and	261	and
3	233	to	160	river	162	river
4	129	river	160	to	148	to
5	100	girl	110	girl	91	help
6	87	help	75	help	69	girl
7	77	into	72	for	68	for
8	74	boys	57	fishing	59	at
9	72	at	57	into	50	they
10	60	for	52	at	49	in
11	60	girls	50	after	47	a
12	59	riverside	46	boys	46	go
13	55	they	44	one	45	fishing
14	47	one	42	a	45	three
15	45	were	42	Siti	42	flower
16	44	was	42	that	39	after
17	43	fishing	41	Is	39	into
18	42	her	41	was	37	hospital
19	37	shocked	39	fall	37	save
20	35	after	38	because	36	fall

4.3 Analysis of the English article *a*: Distribution of pattern across periods of language development

Apart from examination on the frequency of use of the English article *a*, analysis on language development of second language learner in this study also takes into account the patterns of use the selected English article into analysis. This subsection will present the distribution of pattern of use of the English article *a* across the learner corpus. Table 4.2 shows the distribution of the pattern of use of the English article *a* across the three sets of data.

Table 4. 2 Distribution of Pattern of Use of the English article *a* in the Corpus

Patterns of use	Time 1		Time 2		Time 3	
	Frequency	%	Frequency	%	Frequency	%
a+n	11	61	72	90	27	58
a+adj+n	5	27	6	8	11	23
a+v	2	11	2	2	9	19
TOTAL	18	100	80	100	47	100

As can be seen from the table, the type of patterns of use has shown no change over time. It was observed that the second language learners incorporated three main patterns of the English article *a* in their writing throughout the whole learner corpus. The first and the most recurring pattern of use of the English article *a* in the second language learner corpus is the most basic pattern of *a+n*, whereby the English article *a* is used directly with a noun, for instance, *the river*. The second pattern of use is the extended pattern of *a+adj+n*. In this case, it was observed that the second language learners use the English article *a* in an adjective phrase, for instance, *a beautiful flower*. Lastly, the analysis also revealed that the English article *a* was also used in an uncommon pattern in

which it is directly associated with a verb, forming a pattern of *a+v*, for instance *a plucking*.

In contrast to the pattern of use of the English article *a*, the frequency of use of each patterns on the other hand has been evidenced to increase across the three sets of data except for the basic pattern of *a+n*. The pattern of *a+n* was observed to increase in the frequency of use from Time 1 to Time 2, as in 11 times of use and it has increased to 72 times in Time 2, however, the use of the pattern *a+n* in Time 3 has decreased to 27 times only. The observed performance revealed that the English article *a* had been used in three patterns across the three sets of data. The analysis of Time 1 of the learner corpus data of the English article *a* has shown the English article *a* was observed to be used for 18 times in the learner corpus. It was also discovered that English article *a* has been used in three different patterns which are *a+n, a+adj+n* and *a+v*. The highest frequency of the pattern of *a* used in the corpus is the pattern of *a+n* **(61%),** followed by the pattern of *a+adj+n* **(28%)** and lastly an unusual pattern of *a+v* **(11%).**

At Time 2, the English article *a* was observed to be used for 42 times in the corpus (see Table 4.1). The same pattern of use as Time 1 was observed with the highest frequency of the pattern of *a* used in the corpus is the pattern of *a+n* **(90%),** followed by the pattern of *a+adj+n* **(8%)** and *a+v* **(2%).** Lastly, the analysis of Time 3 revealed that the English article *a* occurred 47 times in the corpus (see Table 4.1). It was discovered that English article *a* was used the same manner as in Time 1 and Time 2. The highest frequency of the pattern of *a* used in the corpus is the pattern of *a+n* **(58%),** followed by the pattern of *a+adj+n* **(23%)** and *a+v* **(19%).**

From the table, it can be seen that the pattern of *a+n* was used at the highest frequency at Time 2 which was 72 times (90%) and lowest at Time 1, 11 times (61%). For the pattern of *a+adj+n*, it can be seen that it was used at the highest frequency at Time 3 which was 11 times (23%) and the lowest frequency at Time 1 which was 5 times (27%).

Last but not least, for the uncommon pattern of *a+v*, it was recorded the highest at Time 3 as it was used for 9 times (19%) at lowest at Time 2 as it was used for 2 times (2%). From the analysis, it can be said that language learners use the English article *a* in a more complex manner over time across data collection period.

Through the analysis of the learner corpus, it was observed that the English article *a* was used in the pattern of *a+n* when it is directly followed by a noun group. This pattern is observed as the most recurring pattern of use across the data. Examples from the data include:

(1) At that *a* time, Nadia and Riena are plucking flowers at the riverside. (128a)

A doctor checked Ayda. (125b)

They heard *a* girl asked for help. (102c)

Apart from the most basic pattern of *a+n*, the English article *a* was also observed to occur in the pattern *a+adj+n* in which the English article *a* was used preceding an adjective that modifies the noun following it.

(2) Have *a* shady tree at the river (101a)

After *a* few hour, the girl waked up… (102b)

We looked *a* big tree at the behind big tree (118c)

Apart from the aforementioned patterns of use, it was also observed that learners used the English article *a* in an unusual pattern of *a+v.* The English article *a* was used preceding a verb. Examples from the data include:

(3) Amirul and Hafiz *a* walked at the riverside (109a)

Ahmad, Ali and Abu *a* talk go to the riverside. (118b)

A reached at the hospital the doctor checked Aminah. (105c)

4.4 Analysis of the English article *a*: Analysis of senses across periods of language development

One of the use of collocational information is to highlight the different meanings that a word has (Hunston, 2002). The senses of the meaning of words can be identified by the sequences of morphemes that they normally occur with (Hunston & Francis, 2000). Moreover, the idiom principle states that the meaning of a word is usually derived from the group of words it comes together with (Sinclair,1991). In this study, in order to identify the different meanings or senses in the language system that the English article *a* holds in the learner corpus, the concordance lines were examined and the meanings associated were determined using Collin Cobuild Advanced Learner's Dictionary (8ᵗʰ Edition,2014) and the Oxford Advanced Learners' Dictionary (8ᵗʰ Edition, 2013). The process of analysing the senses of use of the English article *a* was repeated three times and the meanings are cross-checked internally in order to identify the meaning of each instance. Results of the analysis revealed that the English article *a* was used in four different senses across the three sets of data namely as:

1. A determiner when referring to someone or something for the first time. Readers are not aware of which particular person or thing the English article *a* is referring to. It has not been mentioned before and used before countable nouns.

     e.g., Her parents go to the hospital looking for *a* daughter.

2. A determiner when referring to any person or thing in which the writer does not intend to be specific about it.

     e.g., A boy helped the girl.

3. A determiner when referring to quantifier, for instance: a little, a few and a little.

     e.g., A few minutes later, the girl was out from doctor room.

4. A determiner when referring to someone by the job that they have by describing them using their profession group.

      e.g., After they reach, call *a* nurse for help

Based on the findings from the analysis, it was observed that the English article ***a*** is used as a determiner to refer to someone or something for the first time (e.g., *A* big tree nearby the river), the English article ***a*** was used preceding an adjective phrase (big tree). Readers might not know which tree the writer was referring to, it could be any big tree nearby the river. Other examples from the data include:

(1) Siti and Hafizah was happy as *a* lark (101a)

    After that, a doctor checked Murni and gave *a* medicine (123b)

    While carrying *a* pail and fishing rode. (118c)

The English article ***a*** is also used as a determiner when referring to any person or thing which the writer does not intend to be specific about it. Other examples from the data include:

(2) Asmadi was run as *a* lighting to save the girl (105a)

    Suddenly, *a* girls fall into the river want try to pick the flower. (115b)

    Siti and Aminah have *a* trip (102c)

The English article ***a*** is also used as a determiner when referring to quantifier, for instance: a little, a few and a little. Other examples from the data include:

(3) After *a* few minutes later, three boys shocked because they heard the screen. (107a)

    After *a* few minute, Nizam, Nasrul and Syahwal heard the girl screamed. (107b)

    After *a* few minute the girl can go home. (102c)

The analysis also revealed that the English article *a* is also used as a determiner when referring to someone by the job that they have by describing them using their profession group. Other examples from the data include:

(4) After that, *a* doctor checked Murni gave a medicine (123b)

A doctor checked Ayda and gave medicine (125b)

After they reach, call *a* nurse for help (106c)

Initially at Time 1, learners were found to make meanings by using only the first, second and third senses in their writing. However, beginning Time 2, the fourth sense was found to be used in their writing alongside with the other three senses. It was observed that learners used these senses consistently until Time 3.

4.5 Analysis of the English article *the*: Frequency of use across periods of language development

This subsection will present and discuss the analysis of the second English articles concerned in this study which is the English articles *the*. In doing so, the learner corpus, CoMENT was run through the concordance software, AntConc to generate the frequency list as shown in Table 4.2. As can be observed through the table, the frequency of English article *the* was found to decline across the three data collection sessions.

Table 4. 3 Distribution of Frequency of Use of the English article *the* in the Corpus

Function Word	TIME 1		TIME 2		TIME 3	
	Frequency	%	Frequency	%	Frequency	%
The	509	11.5	409	10.6	340	8.74

During the beginning of the data collection session, the English article *the* was recorded to be used for 509 times (11.5%) throughout the learner corpus. However, as the time goes by, it was reported that the frequency of use of the English article *the* has experienced a gradual decrease, as it decreased to 409 times (10.6%) in Time 2 and the frequency went down to 340 times (8.74%) during the third data collection.

Table 4.3.1 below illustrates the rank of the English article the across the whole learner corpus. It can be clearly discerned from Table 4.3 that the English article *the* has been on top of the word frequency list for all three times across the development period. It shows the rank order of the English article *the* which is at the top of frequency list for each data collection. As can be seen, the highest frequency of use of English article was 509 times in Time 1, which then reduced to 406 times in Time 2 and it was calculated to be used 334 times in data of Time 3.

Table 4.3. 1 Comparison of Word Frequency of the Top 20 Words in the Corpus

Rank Order	Time 1 November 2015		Time 2 February 2016		Time 3 May 2016	
1	509	the	409	the	340	the
2	270	and	231	and	261	and
3	233	to	160	river	162	river
4	129	river	160	to	148	to
5	100	girl	110	girl	91	help
6	87	help	75	help	69	girl
7	77	into	72	for	68	for
8	74	boys	57	fishing	59	at
9	72	at	57	into	50	they
10	60	for	52	at	49	in
11	60	girls	50	after	47	a
12	59	riverside	46	boys	46	go
13	55	they	44	one	45	fishing
14	47	one	42	a	45	three
15	45	were	42	Siti	42	flower
16	44	was	42	that	39	after
17	43	fishing	41	is	39	into
18	42	her	41	was	37	hospital
19	37	shocked	39	fall	37	save
20	35	after	38	because	36	fall

This consistency on the frequency of use of the English article *the* in the learner corpus is on par with the claim made by Sinclair (1991) that the English article *the* is the most frequent word in a corpus of 20 million words. In this learner corpus, despite the fact that the size of the corpus is smaller, the frequency of the English article *the* still reflects the notion that it is the most recurring free morpheme in English (Master, 1997).

4.6 Analysis of the English article *the*: The distribution of pattern across periods of language development

Apart from investigation on the frequency of use of the English article *the*, this study also looks thoroughly at the patterns of use of the selected English article. This subsection will present the distribution of the pattern of use of the English article *the* across the learner corpus. Table 4.4 below summarises the distribution of the pattern of use of the English article *the* across the three sets of data.

Table 4. 4 Distribution of Pattern of Use of the English article *the* in the Corpus

Patterns of use	Time 1		Time 2		Time 3	
	Frequency	%	Frequency	%	Frequency	%
the+n	460	90.3	366	90.1	305	91.3
the+adj+n	37	7.3	37	9.1	22	6.6
the+v	12	2.4	3	0.8	7	2.1
Total	509	100	406	100	334	100

The observed performance revealed that the English article *the* had been used in three patterns across the three sets of data. Analysis of Time 1 of the learner corpus data with keyword *the* has shown that the English article *the* occurred for 509 times in the corpus. It was discovered that English article *the* was used in three different patterns which are *the+n, the+adj+n* and *the+v*. The highest frequency of the pattern of *the* used

48

in the corpus is the pattern of *the+n* **(90.3%),** followed by the pattern of *the+adj+n* **(7.3%)** and lastly an uncommon pattern of *the+v* **(2.4%).**

Analysis of the learner corpus of the data collected at Time 2 revealed that the English article *the* was observed to occur for 406 times in the learner corpus. The highest frequency of the pattern of *the* used in the corpus is still the basic pattern of which the English article *the* is directly followed with a noun group which forms the pattern of *the+n* **(90.1%).** This is followed by an extended pattern of *the+adj+n* **(9.1%)** whereby the English article *the* is used in an adjective, and lastly the pattern of use that recorded the least number of frequency which is the uncommon way to incorporate the English articles in writing, in which the English article or the English articles is directly associated with a verb forming the pattern of *the+v* **(0.8%).**

The analysis of the learner corpus at Time 3 observed that the English article *the* occurred for 334 times throughout the learner corpus. It was also observed that the English article *the* was used in the same three patterns which are *the+n, the+adj+n* and *the+v*. The highest frequency of the pattern of *the* used in the corpus is the pattern of *the+n* **(91.3%),** followed by the pattern of *the+adj+n* **(6.6%)** and *the+v* **(2.1%).**

From the table, it can be seen that the pattern of *the+n* was used the most at Time 3 with 91.3% and was used the least at Time 2 with the percentage of 90.1%. For the pattern of *the+adj+n*, it can be seen to be used the most at Time 2 at 9.1% and the least used at Time 3 at 6.6%. The uncommon pattern of *the+v* was recorded to be used the most at Time 1 at 2.4% (12 times) and the lowest frequency of use at Time 2 which was 0.8% (3 times). It can be said that over time, language learners showed increase and decrease in the language pattern of the English article *the*.

The most frequent recurring pattern of use of the English article *the* is the most basic pattern which is *the+n.* This pattern occurs when the English article *the* is directly

followed by a noun group. This pattern is observed as the most recurring pattern of use across the data. Examples from the data include:

(1) *The* doctor checked Aminah. (126a)

The girl is happy. (138b)

Suddenly, Siti fall into *the* river. (101c)

Apart from the most basic pattern of ***the+n,*** the use of an extended pattern of ***the+adj+n*** was also observed to occur across the data of the learner corpus. The English article ***the*** was observed to occur preceding an adjective which modifies the noun following it. Examples from the data include:

(2) …jumped into the river to save *the* beautiful girl. (107a)

Siti saw *the* beautiful flowers (102b)

The three boys heard a girl help. (141c)

On the other hand, the analysis also observed the use of the English article ***the*** in an unusual pattern of **the+v**. This pattern occurs when the English article ***the*** is directly followed by a verb. Examples from the data include:

(3) Imran and Shafid heard *the* screamed. (127a)

Fakri and friend were walking near the river for *the* fishing. (125b)

We go to *the* walking at the behind home. (118c)

In the conclusion, result from this subsection reflected that in comparison with the distribution of pattern of use of English article ***a,*** the distribution of pattern of use of the English article ***the*** indicated some distinct changes in the patterns of use. However, in contrast to the frequency of use of the pattern for the English article ***a***, the frequency of use of the English article ***the*** was observed to decrease across the three sets of data.

4.7 Analysis of the English article *the*: Analysis of senses across periods of language development

A thorough examination on the selected English articles in the concordance lines is crucial in order to identify the various meanings that they hold as one of the use of collocational information is to highlight the different meanings that a word has (Hunston, 2002). Similar to the analysis of the English article *a*, the Collin Cobuild Advanced Learner's Dictionary (8th Edition, 2014) and the Oxford Advanced Learners' Dictionary (8th Edition, 2013) were used in determining the meanings associated with the English article *the* in the COMENT. Analysis of senses was repeated three times in order to identify the meaning of each instance.

The results of the analysis revealed that the English article *the* was used in four different senses across the three sets of data namely as:

1. A determiner to refer to someone or something that have already been mentioned or identified. It is used at the beginning of noun groups.

    e.g., Aidil, Hafiz and Zaid heard *the* scream.

2. A determiner to refer to something in our general experience of the world. It means that the English article *the* is used to describe things that are associated with the nature of the word, for example, the sun.

    e.g., Anna help me to plucked *the* flowers.

3. A determiner to refer to people, things, services or institutions that are associated with everyday life. It is used in front of nouns. In this sense, we understand that

the English article *the* is used to describe nouns that are associated with our everyday life, for example, the doctor.

e.g., They brought *the* girl to *the* hospital.

4. A determiner to refer to titles, name of place or other names, for instance, The Royal Albert Hall.

e.g., The girl was saved at *the* Hospital Sultan Haji Ahmad Shah.

Based on findings from the analysis, it was observed that the English article ***the*** is used at the beginning of noun groups as determiner to refer to someone or something that has already been mentioned or identified. It is used at the beginning of noun groups. Other examples in the data include:

(5) Farhana say thank you to *the* boy for saving her. (122a)

After ten minutes, they sent *the* girl to the emergency room. (107b)

Mamat jumped into *the* river. (102c)

The analysis also revealed that the English article ***the*** is also used as a determiner when referring to something in our general experience of the world. It means that the English article ***the*** is used to describe things that are associated with the nature of the world, for example, the sun. Other examples in the data include:

(6) The girls are plucking flowers at *the* riverside. (134a)

Then, Siti saw *the* beautiful flowers. (102b)

Siti and Zara pluck *the* flowers beside *the* river. (108c)

Apart from that, it was also discovered that the English article ***the*** is used as a determiner to refer to people, things, services or institutions that are associated with

everyday life. It is used in front of nouns. In this sense, we understand that the English article *the* is used to describe nouns that are associated with our everyday life, for example, the doctor.

(7) *The* doctor check Aminah. (126a)

The doctor checked a girl and she gived medicine. (107b)

After 5 minutes, ambulance took Athyrah into *the* hospital. (128c)

The analysis on the concordance lines of the learner corpus also revealed that *the* English article *the* is used as a determiner to refer to titles, name of place or other names, for instance, The Royal Albert Hall. Other examples in the data include:

(8) Doctor checked to *the* Laila (129a)

The ambulance came and took Ayda to *the* Kajang Medical Centre. (125b)

We carrying Syakirah go to *the* hospital Sultan Ahmad Shah. (118c)

Lastly, based from the analysis of the English article *the* in this learner corpus, it could be discerned that there is no obvious change tracked through the way learners use the English article *the* in their writings. It is also indicated that the second language learners in this study also showed consistent performances in their expression of senses as evidenced through the texts.

4.8 Conclusion

This study looks at how second language learners develop in their use of English articles with focus on the frequency of use, distribution of pattern as well as senses in context of the English articles *a* and *the* or known as the English articles system over time

in a longitudinal corpus. Based on the analysis conducted on the concordance lines, there are a few findings that have emerged from the analysis of the CoMENT in this study. Over time, it was observed that changes were tracked in the corpus from a few aspects which are the changes in frequency of use of the selected English articles, distribution of patterns of use as well as their meaning making in context.

In the CoMENT, changes were evidenced through the way leaners used the English articles *a* and *the* in their writings. It was revealed that the English article *a* shows increase in frequency of use over time. The learners were found to use the English article *a* more frequently in the learner corpus over time. However, in terms of senses or meaning making, it was evidenced through the concordance lines that the learners used consistent senses over time. The learners incorporated more of the English article *a* in their writing, but, they concurrently produced the same meaning of the English article *a*.

In contrast with the English article *a,* the English article *the* on the other hand is observed to decrease in frequency of use. However, contrary to its frequency, the learners were observed to incorporate more meaning associated with the English article *the* in their writings. It was evidenced through the analysis of the concordance lines that the learners used more variation of meanings for the function *the*, as compared to the English article *a* in their writing across the three points in time.

This observation reflects the complex linguistics system that bi/multilingual learners have. Learners were observed to incorporate more meanings to the English article *the* as compared to the English article *a*. However, in terms of the frequency of use, learners were observed to incorporate the English article *a* more than the English article *the.* The process of second language development in their mind is unpredictable, suggesting a chaotic pattern of development (Larsen-Freeman, 1997, 2006). This was reflected through the analysis of the English articles *a* and *the* in the learner corpus*.* The frequency of use, the pattern of use as well as the senses incorporated in the writings were

54

in a waxing and waning situation whereby one cannot simply predict what will come next as it differs from each aspect.

CHAPTER 5: DISCUSSION AND CONCLUSION

This chapter will present the discussions based on the findings obtained in the analysis conducted in Chapter 4. This section will be organised by subsections whereby each subsection will represent and discuss the research questions posed earlier in Chapter 1.

5.1 Introduction

Over time, changes are tracked in the corpus from a few aspects which are the changes in frequency of use of the selected English articles, distribution of patterns of use as well as their meaning making in context. This chapter presents the discussion of the result of the analysis presented in the previous chapter.

Through the analysis of the learner corpus, the CoMENT, it was discerned that out of the three English articles *a, an* and *the*, only two were found to be used across the corpus. There was no evidence of the use of the English articles *an* across the CoMENT. This study specifically analysed a few aspects which can indicate the second language development of the second language learners.

One of the aspects being analysed in this study is the frequency of use of the English articles. Throughout the learner corpus, it was revealed that frequency of use of the English article *a* experienced some growth over time. A developmental pattern can be observed in both English articles *a* and *the*. The second language learners were observed to incorporate more use of the English article *a* in their writing. This is evidenced through the frequency table presented in Chapter 4. Based from the table, it could be seen that the second language learners in the corpus have used the English articles *a* in their writing more frequently as time goes by. However, in contrast to the frequency of use of the

English article *a*, the English article *the* was seen to gradually decline through the three times of the data collection. At the beginning of the data collection, the second language learners were found to use a massive amount of the English article *the* in their writing. But as the time goes by, the frequency of use of the English article *the* has gradually experienced reduction. This illustrates the concept that the second language learners use less of the English article *the* in their writing as they progress in their language learning.

Secondly, this study was also concerned about the pattern of use of the English article system. Patterns and meaning are connected (Hunston, 2000). Findings from the analysis revealed that the distributions of the patterns of use of English articles *a* and *the* showed no obvious change over time. Both English articles are used in three different patterns consistently across the data which the basic pattern of *a+n* and *the+n,* the extended pattern of *a+adj+n* and *the+adj+n*. The patterns observed for both of the English articles showed no obvious change across the three points in time. This indicates that the second language learners use the English articles in a consistent and stable manner as compared to their frequency of use. However, it was also discovered that the English articles are used in an unusual pattern of *a+v* and *the+v.*

This situation might be explained through the perspective of transfer of language in second language learning. The second language learners are prone to undergo transfer in their process of language learning. A better explanation at this is the concept of translanguaging by Velasco and Garcia (2014) in which they claim that inside the mind of a bi/multilingual, there will be a unique process to create a meaningful communication using two or more linguistics repertoire. The bi/multilingual minds possess a very spectacular and distinctive language system equipped with their own grammar sets in which it may contain variation from the standard language.

This study also looks at the way the second language learners incorporate the use of English articles in their writing from the perspective of meaning making. Patterns and

meaning are closely associated. Changes in the way second language learners associated with the use of English articles in their writing may reflect their development in second language learning and acquisition. In the case of the English articles *a*, it was initially observed that the second language learners were found to use only three different senses in their writing. It was evidenced through the text that the learners only used the English articles to convey these three meanings;

1) Someone or something for the first time. Readers may not be aware of which particular person or thing that the English article is referring to.

2) Any person or thing in which the writer does not intend to be specific.

3) Quantifiers such as a little, a few and a little.

However, beginning time 2, the second language learners were observed to use another type of senses which is used to refer to:

4) Someone by the job that they have.

However, the analysis of the English article *the* revealed that there is no obvious change that was tracked through the way learners use the English article *the* in their writings. The students also showed consistent performances in the expression of senses as evidenced through the texts. Throughout the analysis of the concordance lines, it was evidenced through the text that the second language learners incorporated these four different senses to convey intended meaning in their writing which are:

1) Someone or something that have already been mentioned or identified.

2) Something in our general experience of the world

3) People, things, services or institutions that are associated

4) Titles, name of place or other proper names.

58

Henceforth, it could be said from this study that the use of English article system across the three points in time has undergone a mix of increasing and decreasing pattern of use. This situation reflects the concept that the second language learners are still expanding their linguistic repertoire and as time goes by their linguistic ability will continue to grow and expand. The waxing and waning of the language properties illustrates the inner structure of the language development that occurs in the mind of a bi/multilingual. This is supported by Chau (2012) where he claims that language is not a static object; it will undergo changes, growth and organizes itself.

Based on the analysis, it was noted that throughout the three points in time, there are changes in a few aspects can be observed over time. This chapter attempts to address the research questions initially raised in Chapter 1 which are:

1. What are the most frequent patterns of use of the article system at Time 1 of the learner data?

2. Is there evidence for changes in the use of the article system across the three points in time in the learner data?

How might such differences be explained in terms of language development?

5.2 Research Question 1

What are the most frequent patterns of use of the article system at Time 1 of the learner data?

This research question investigates the most frequent patterns of use of the English article at time 1 of the CoMENT. The analysis of the data has observed that there are two most frequent patterns of use of the article system at Time 1 of the learners' data. The most frequent patterns of use of the English article system at Time 1 of the CoMENT is the utmost basic pattern of *a+n* for the English article *a* and pattern of *the+n* for the English article *the*. There was no evidence of the use of English article *an* that fits into the criterion outlined earlier in Chapter 1 throughout the analysis of the concordance lines. This pattern occurs when the English articles *a* and *the* are directly followed by the noun group.

Table 5. 1 The Pattern of Use of the English Article System at Time 1 of the CoMENT

Patterns of use	A		THE	
	Frequency of use	Relative frequency	Frequency of use	Relative frequency
art + n	11	0.314	460	13.143
art + adj + n	5	0.143	37	1.057
art + v	2	0.057	12	0.343

Table 5.1 illustrates the frequency of patterns of use of the English article system *a* and *the* as observed and discerned from the corpus. As evidenced through the analysis of the concordance lines, the most recurring patterns of use of the aforementioned English articles at the beginning of the data collection session was the most basic pattern of use which are *a+n* and *the+n*. The pattern of *a+n* was observed to occur for 11 times in the

learner corpus while the pattern of *the+n* on the other hand was recorded to be used for 460 times.

Patterns were encoded by a mapping of the word-classes involving a pattern on to the functional label by Hunston (2000). In this study, it is observed that the most frequent pattern that emerged in Time 1 is the pattern of which the English article *a* is directly followed by a noun group. The same case applies to the English article *the* whereby the most frequent pattern that was observed in the second language learners data was the pattern of *the+n* in which the English articles *the* is directly used before a noun group. The patterns of *a+n* and *the+n* in the concordance lines includes examples in which the noun is preceded by an article.

(1) At that *a* time, Nadia and Riena are plucking flowers at the riverside. (128a)

   A doctor checked Ayda. (125b)

   They heard *a* girl asked for help. (102c)

(2) *The* doctor checked Aminah. (126a)

   The girl is happy. (138b)

   Suddenly, Siti fall into *the* river. (101c)

As explained in Chapter 3, Time 1 of the data collection session serves as the starting point of this study. It is not by any means reflecting the total state of the second language learners, but instead it serves as the point to start monitoring the changes and development that takes place in the learner corpus.

5.3 Research Question 2

Is there evidence for changes in the use of the article system across the three points in time in the learner data?

The second research question looks at the changes in the use of the English article system across the three points in time in the second language learners' data and concerns as to what extent the use of the article system differs across the three points in time. Based on the analysis, changes were evidenced in terms of the frequency of use of the English articles, the patterns of use of the English articles as well as the senses associated with the English articles.

In terms of the frequency of use of the English articles, the English article *a* was found to increase over time, however, the English article *the* was found to experience a decline over time in the learners' data. Table 5.2 presents the frequency of use of the English article system of *a* and *the* in the corpus.

Table 5.2 presents the frequency of use for both of the English articles *a* and *the* across the three times of data collection in the CoMENT. It can be clearly observed that the frequency of use for the English articles *a* has experienced a gradual increase beginning Time 1 until Time 3 of the data collection. However, in comparison with the frequency of the English article *a*, the English article *the* was observed to undergo a gradual decrease from Time 1 of the data collection until Time 3. These observations are also evidenced through the increase in the calculated percentage of use for both of English articles *a* and *the*.

Table 5. 2 The Frequency of Use for *a* and *the* across the CoMENT

WORDS	TIME 1		TIME 2		TIME 3	
	Frequency	%	Frequency	%	Frequency	%
A	18	0.4	42	1.1	47	1.21
The	509	11.5	409	10.6	340	8.74

In terms of the patterns of use, findings revealed that the distributions of the patterns of use of English articles *a* and *the* showed no obvious change over time. Both English articles are used in three different patterns consistently across the data which the basic patterns of *a+n* and *the+n*, the extended patterns of *a+adj+n* and *the+adj+n*. It was also discovered that the English articles are used in an unusual patterns of *a+v* and *the+v*. Table 5.3 presents the patterns of use of the English articles *a* and *the* across the CoMENT.

Table 5. 3 Distribution of Patterns of Use of the English Articles *a* and *the* across the CoMENT.

Patterns of use	Time 1		Time 2		Time 3	
	Frequency	%	Frequency	%	Frequency	%
a+n	11	61	32	76.2	27	57.4
a+adj+n	5	28	6	14.3	11	23.4
a+v	2	11	2	4.7	9	19.1
the+n	460	90.3	366	89.5	305	89.7
the+adj+n	37	7.3	37	9.04	22	6.47
the+v	12	2.4	3	0.73	7	0.2

Table 5.3 presents the distribution of patterns of use of the English articles *a* and *the* across the CoMENT. From the table, it can be seen through the table that the patterns of use of both of the English articles did not undergo any changes across the three points of

the study. It can be observed that the second language learners' use of the English articles is consistent as they use it in three fixed patterns, whereby the English articles are:

1) Directly associated with a noun group.

    a. In this pattern, the English articles *a* and *the* are observed to be used directly with a noun group, forming a noun phrase, for instance, *the* river, *the* flower, *a* boy and *a* girl.

2) Used in an adjective phrase whereby the English articles is used in an adjective phrase

    a. In this pattern, the English articles *a* and *the* are observed to be used directly with an adjective phrase, for instance, *the* beautiful girl and *a* beautiful flower.

3) An uncommon pattern of which the English articles is directly associated with a verb.

    a. Even though this pattern is quite uncommon, it was observed throughout the learner corpus, the second language learners tend to consistently incorporate this pattern in their writing. In this pattern, the English articles will be directly followed by a verb, for instance, *a* plucking, *the* plucking.

In terms of the senses, changes were evidenced in the way learners use the English article *a.* Table 5.4 represents the use of different senses of the English articles *a* across the three points in the learner corpus, the CoMENT.

Table 5. 4 The Use of Different Senses of the English Articles *a* across the CoMENT

	Time 1	Time 2	Time 3
Sense 1 Refers to someone or something for the first time.	/	/	/
Sense 2 Refers to any person or thing in which the writer does not intend to be specific about it.	/	/	/
Sense 3 Refers to quantifier, for instance: a little, a few and a little.	/	/	/
Sense 4 Refers to someone by the job that they have by describing them using their profession group.	-	/	/

As can be shown from Table 5.4, initially at Time 1, the second language learners were found to make meanings by using only the first, second and third senses in their writing. However, beginning Time 2, the second language learners were observed to use the fourth sense in their writing alongside with the other three senses. It was observed that learners used these senses consistently until Time 3. However, the analysis of the English article *the* revealed that there is no obvious change was tracked through the way learners use the English article *the* in their writings. The students also showed consistent performances in the expression of senses as evidenced through the texts. Table 5.5 represents the use of different senses of the throughout the learner corpus.

Table 5. 5 The Use of Different Senses of the English Articles *the* across the CoMENT

	Time 1	Time 2	Time 3
Sense 1 Refers to someone or something that have already been mentioned or identified.	/	/	/
Sense 2 Refers to something in our general experience of the world.	/	/	/
Sense 3 Refers to people, things, services or institutions that are associated with everyday life. It is used in front of nouns.	/	/	/
Sense 4 Refers to people, things, services or institutions that are associated with everyday life.	/	/	/

It could be seen that students showed consistent performance in the use of the English articles of *a* and *the*. Through observation from Time 1 to Time 3, students showed changes in the way they associate the English articles *a* and *the* in their writing. They progressed through the years, as they were observed to associate more senses of the English articles *a* and *the* in their writings.

5.4 Research Question 3

How might such differences be explained in terms of language development?

This research question looks at the significance of the differences discussed in research question 3 and how the differences are explained with regard to language development. It was reflected in the analysis that the use of English article system was not fixed and it did not seem to follow any order of development. Instead, the second language development evidenced through the frequency of use and the pattern of use of the English article system tend to be in a mixture of increasing and decreasing pattern. However, in contrast to the frequency and patterns of use, the senses or the meaning making involving the English articles, it could be discerned that the participants used

66

more variety of the senses towards the end of the data collection session. It could be safely understood that as time goes by, the second language learners will begin to use lexis with an extensive variety of meanings, and in this matter, the second language learners' use of English articles has undergone an expansion over time.

Based from the evidences gathered from this study, it is clearly reflected and understood that second language learners' experience of the second language development is rather chaotic and unpredictable, which suggests a chaotic pattern of development (Larsen-Freeman, 1997, 2006) of the English articles. The process of learning a second language itself requires the learners to process, construct and acquire the content of the language and assimilate the content with the existing linguistic system in their minds (Velasco & Garcia, 2014). Second language learners will own at least two different linguistics repertoire and these two languages (or more) might be separated or closely linked in the mind of a bi/multilingual, but eventually one language may affect the other and consequently second language learners will create their own linguistic system with their unique properties (Cook, 2003).

This concept is parallel with the notion of multicompetence in which he describes it as "the compound state of a mind with two grammars" (Cook, 1992 pg. 557). As language practitioners, it is equally important to hold on to the principal of treating the second language learners in their own right without any comparison with the native speakers' production norm. This will in turn creates the concept of not viewing them as deficient language users, but rather, the second language learners will now be considered as different and will be viewed as a total distinction from native speakers.

This situation is also supported by the idea of translanguaging in which it looks at how multiple languages relate to each other inside the mind of one person, during both process of acquiring the second language and while using it (Gracia, 2014). Instead of focussing on producing the traditionally correct and typical English language norm, the

notion of translanguaging focusses on the flexibility and meaningful actions taken by bi/multilingual language users in order for them to communicate effectively, without viewing the languages of bi/multilinguals as separate linguistic systems (Velasco & Garcia, 2014).

This process provides visions on how two or more languages interact and leave an impact on the learning process. From this view, the bi/multilinguals are put in a situation in which they exist in a functional connection with the other languages that they practise and consequently they will be establishing an integrated system of language learning. Language is not a static object, instead, it grows, changes and organises itself through an organic way (Chau, 2012).

Instead of viewing the language production of bi/multilingual language users as a more or less successful attempts to reproduce the structural properties of the target language (Klein & Perdue, 1997), we may now divert our perspective into viewing their language production as the outcome of the flexible and meaningful actions of language repertoire in the minds of bi/multilingual language users within the dynamic process of making sense of the world.

5.5 Discussion and comparison with other relevant studies

This study is comparable to a prominent study reported by Hakuta (1976) which shows both English articles *a* and *the* are acquired as a system. He studied the order of acquisition of grammatical morphemes including articles in the interlanguage of an ESL child. The study seeks to answer questions pertaining to the order of grammatical morphemes in the interlanguage of a child. However, it was observed that the subject of the study used *the* better than *a*. In the study, Hakuta has observed a 5-year old Japanese girl acquiring English in a natural setting and her natural speech was recorded for over a

period of 60 weeks. Findings from the study concluded that the English articles *a* and *the* are acquired as a system. It was initially shown that the performance of *the* was initially better than *a*.

Other than that, Huebner (1983) conducted another study in the same field. He was among the earliest linguists who looks at the aspect of English articles development. He conducted a case study on a 23-year-old speaker of Hmong, known as Ge, acquiring English in a natural setting by looking at the development of the articles system in an adult's interlanguage. The longitudinal study was conducted for 54 weeks and for every three weeks, a tape was made of the subject's narratives. He adopted the Bickerton's model in conducting the study.

Huebner's initial intention was to look at the development of tenses in Ge's language production, however, upon finishing the data collection, Ge's language production was found to be massively influenced by his changing form of is (a), the English article *the*, as well as the anaphoric devices of pronouns and zero (Huebner, 1983). He shifted and investigated the use of definite article, *the*, by the subject of the study. The result showed that English article *the* emerged early and it also experienced *the*-flooding. While the English article *a* was found to appear later in the subject's L2 acquisition. This situation is similar to Hakuta (1976), the English article *a* was reported to appear later in the speakers' acquisition.

Huebner was reported to examine all instances of the article the used by the subject of the study, not just the one in obligatory contexts and it revealed the range of use of the form. After one year of data collection, he identified six stages in Ge's language acquisition. It was observed that the variation in the use of the English article *the* was acquired and is not random. The initial system used by the subject was neutralised before the subject started employing a new system in his language production. In the case of the

English article *the*, Ge was found to gradually eliminating the 'incorrect' form of the article before incorporating it 'correctly' in his language production.

Reported studies have shown that there is a system to the acquisition of the article system despite its variation. Previous studies have revealed that variation in the use of English articles in systematic in pattern and by agreeing to the statement that both article accuracy and frequency are correlated with the L2 learners' English proficiency, thus results in the variation.

The present study seeks to answer the questions pertaining to the development of language in terms of articles system by the second language learners, with focus on the frequency of use, distribution of pattern as well as the senses. Across the points in the study, it was observed that the use of English articles *a* has experienced an increase in frequency as opposed to *the*. However, learners were analysed to incorporate more meanings related to the senses of *the* in their writing as compared to *a*. In general, it could be understood that acquisition of English articles is inconsistent with one another (Hakuta,1976; Huebner,1983a; Parrish, 1987). Even though the acquisition order may vary across the studies, it is shown that the first acquired article is the definite article *the*. It is revealed that as *the* is observed to be acquired earlier that the other articles, learners also exhibit the habit of *the*-flooding in the process of the language development.

The bi/multilingual language are unique users of language as they are able to produce their own grammar sets with their unique linguistics repertoire which is created out of their own internal processes in response to the second language data received Cook (2008). Henceforth, the process of language production in a bi/multilingual language user mind is more complex as compared to the monolinguals, as they possess two or more linguistics repertoire and those systems are interrelated. The process of producing language in the mind of a bi/multilingual is clearly not a linear process as it emerges from the intertwined linguistic systems in their minds.

This unique feature of language processing in mind is closely linked to the concept of translanguaging and complemented by the idea of multicompetence. Language learners play an important role in the process of language acquisition as language development only occurs when language learners interact and make use of the acquired knowledge of language that they are exposed to, which consequently producing a new language system on its own (Cook, 2008). The central idea in this study is to make an attempt to understand and provide reasonable explanation on how bi/multilingual language users develop in their language production, with focus on the aspect of English article system, through the medium of written English.

5.6 Learners' unique ability to form meanings

The idea of translanguaging has suggested that a bilingual or multilingual language user owns a very unique linguistic system. The system that they have is not separated or different. As we view language as a process or an object (Swain, 2006), the bilingual language users are seen as individuals who are using language as a medium or platform in order for them to engage themselves in a meaningful conversation by making, conveying and understanding meanings. The bilingual language users are reported to perform selective use of the language features in their minds in order to use the language. In the process, the bilingual language users did not use only a system of language in their mind, but rather they constantly switch and select their language repertoire according to their needs to meet the requirements of the task. It aligns with the notion of translanguaging introduced by Garcia (2004).

Second language learners have the ability to understand and interact using two or more languages at the same as being suggested as "two monolingual in one". The process of meaning making is exclusive for the bi/multilingual second language learners as they

incorporate the interaction of two or more languages in order to create meaningful communication (Velasco & Garcia, 2014). Bi/multilingual language learners will undergo the self-regulating mechanism whereby they will be able to deeply connect and engage in the process of language learning rather than merely attending the teaching and learning process. This process is not observable to the naked eye, but the evidence of this process is projected through the bi/multilingual learners' second language use and production of the second language learners.

The concept of translanguaging is a complete distinction from codeswitching. It should be understood that the notion of translanguaging is beyond a mere codeswitching process as the latter is more concern in viewing and treating two languages as a separated system. The fact that the users of language select different variety of language according to the appropriateness of the task proves that not all languages and linguistic variety is equally suitable to all speech events. (Romaine, 2006). As codeswitching requires the speaker to switch off one of the language during interaction and use it when activated. While translanguaging on the other hand promotes the idea that two or more languages that exist in the bi/multilingual minds forming an integrated system and it encourages the interaction between the two or more languages in the mind of a bi/multilinguals in order to them to be able to create meaningful sentences to achieve optimum level of communication.

On the other hand, during the process of teaching and learning of the second language, it is fully encouraged to incorporate the second language learners' first language throughout the process, as it will facilitate the process. By doing so, the target achievement and objectives of the lesson can be achieved more effectively. A second language learner should not be denied his or her right to incorporate the use of their first language in the process of second language learning due to the fact that a bi/multilingual

second language learner is different from a monolingual (Cook, 2013). Hence, they should be treated in their own right.

5.7 Pedagogical Implications

The ultimate goal to language teaching and learning is not to achieve perfect imitation of the target language norm, but rather knowing how students learn (Garcia, 2014). Emergent bi/multilingual language users are prone to face difficulties in manoeuvring their language productions through the process of acquiring a second language. Hence, it is the teachers' responsibility to provide a safe environment for the learning process to take place and encourage the learners' participation with the aim to strengthen their understanding of the content knowledge.

Garcia and Sylvian (2011) and Garcia (2014) agree that learners should be given autonomy to incorporate their first language in a flexible manner during classroom to aid in the process of them communicating effectively. This is in line with Esquinca, Araujo and de la Piedra's view (2014), suggesting that language mediates understanding. The use of the second language learners' first language or mother tongue should not be curbed or abandoned in the second language classrooms. It must be noted that incorporating a pedagogy that appreciates learners' cultures and background to write in an additional language has been consistently helpful (Velasco & Garcia, 2014). This promotes deviation from the notion put forward by Klein and Perdue (1997) who claim that teaching is a normative process and it is the teachers' obligation to bring the learners as close to the norm as possible, thus negating the notion of errors as the result of interference from the L1 in the learning process of a second language and eliminates the idea that errors or 'incorrect' utterances of the target language are the evidence of the

acquisition process by Corder (1967). He sees the making of errors as a strategy from the second language learners and it proves the existence of the learner-internal processing.

A language learning situation that provides support and inspires students to try and use whatever knowledge and second language skills that they have in order to communicate effectively in second language must be very accommodating than the one which dwells on the errors, corrections and a failure to be perfectly accurate kind of language learning (Cook, 2003). Hunston (2002) posits the idea that second language learners should be given the autonomy to approach the language in the most comfortable way for them. They should not be in any form of stress and forced to learn a second language via lexical chunks exclusively, for sometimes second language learners will prefer to learn via grammatical rules and lists of lexical items.

Studies on learner corpora have opened up the space for such approach. Indeed, a learner who is willing to guess, take the risk of making mistakes are the one who will tend to be more successful as compared to the passive learner, but what is the point of comparing the second language learners' production of language to the norm of native speakers if it is obvious that they are completely different and possess unique characteristics as compared to the monolinguals? It is very essential to treat the second language learners' production of target language in their own right (Cook, 1991) without putting them into any comparison with the target language norm or the native speakers' production norm.

Studies on learner corpora has shed light on the way learners of English as second language use the language by providing evidences about the process of language learning and thus contributes the field of second language acquisition (SLA) (McEnery & Gabrielatos, 2006). On the other hand, this study can benefit teachers by revealing a clear description as to how their students use English articles in their writing and it reflects the progress in their acquisition in the area. Asmeza, Noor Hayati and Norwati (2013) in their

study concerning errors committed by learners in the use of preposition *in* and *on* have concluded that corpus based studies can provide such insights in the area of second language acquisition studies and this will indeed help teachers to have a better understanding of how their students learn and acquire the English articles. This is supported by Sinclair (1991) who posits that through the means of learner corpora, we are able to study a large text corpus in a systematic manner in which we will have the access to a quality of authentic evidence of the learner language which have not been available before. Through the means of learner corpora, every parts of the language production can be thoroughly investigated and it will reveal the description of the characteristic of the learner language.

In this study, evidence discerned from the learner corpus has revealed the patterns of language use by the learners' and will be useful for the pedagogy and the teachers. Corpus-based research of learner language contributes to English language teaching by aiding in the language instruction and the process to develop and compile the pedagogical and reference materials for different level of competence. The study on learner language will also shed light on the process of language learning (McEnery & Gabrielatos, 2006).

Consequently, it will assist teachers in planning and developing a more innovative and practical lesson plan for the teaching and learning process. This will in turn leads teachers to develop better teaching materials in teaching and learning process. The traditional role of a teacher will be expanded and enhanced as the use of corpora in language teaching will offer space for learner independence (McEnery & Gabrielatos, 2006). This will help the second language learners to facilitate their understanding on the English language, hence improving their language use.

5.8 Recommendation for future studies

As being mentioned in Chapter 1, this study is intended to cover the scope within the context of the selected school, mainly because it was conducted on the students of the selected school only. Owing to the size of the learner corpus, the results of this study are not conclusive. The CoMENT did not contain any instances of the use of the article *an*. This could be due to the size of corpus itself. It would be ideal and fruitful if the future research to be conducted in the same area is done based on a larger scale with a wider selection of participants.

It is important to note that this homogenous learner corpus was collected from the writings of narrative recounts from a set of lower secondary students. The students were in the average of 13 years old when the data was first collected and they were 14 years old when the data collection ended. Different set of students will tend to project different ways and sophistication in the use of their second language. It is advisable and wise for the future research to build a learner corpus by gathering data from different age groups, for instance, the upper secondary students or from the tertiary level students.

One interesting fact about Huebner's (1983) study is that, it is pointed out that the differences in approaching the data analysis will lead to different outcome concerning the nature of interlanguage. I believe that the findings from such study will provide more insights on the language use among second language learners with various features of language production.

5.9 Limitations

It would be challenging for a masters' dissertation to discuss everything pertaining to the language development of second language, through the lens of the English article

system. Although the researcher has tried her best to ensure the quality of the research at its best, there are a few limitations that have to be pointed out. Firstly, this study was only conducted for five months. A longer time frame would benefit the research through observation of the patterns of use of the English articles system, hence, giving a fuller picture on the language development among second language learners.

Last but not least, the limitation of this study is on the representativeness of the corpus. As being mentioned in Chapter 3, the CoMENT was built from data collected from only one school in Malaysia, specifically in Temerloh, Pahang. Participation from various school in Malaysia would give a better and holistic view of language development among second language learners in Malaysia.

5.10   Conclusion

This study aimed to examine the development of language among language learners based on the use of the English article system, as well as to add on the existing literature pertaining to language development in general and the English article system in particular. At this point, I believe that this study has identified the changes in the way learners use the English article system over time.

Nonetheless, the discussion provided on the language development in second language learners always has something new to offer. There is so much potential in the field of corpus-based study. Insights from studies in the same field will shed lights on the field of corpus linguistics and second language acquisition.

There is so much more of a second language use rather than the errors. The arguments imply that errors are not everything that one can see in the second language learners' language production and we have also explained that a bi/multilingual learner is different, a total distinction from a monolingual. The fact is simply because they own

more than one grammar systems in the linguistic repertoire and the ability to incorporate all the languages in their daily life is tremendously great.

A learner language is always in the process of development and formation (Tarone & Sweirzbin, 2009). The variations that emerge in a learner language and the target language norm does not mark the deficiency of the second language learner, instead, it is the proof of the multi-linguistics features existing in the mind of the second language learners and the evidence of the translanguaging process. These "errors" are the proof of the language learning process and the definite separate system of language that a language learner use (Corder, 1967).

Language is dynamic and it is everly changing according to the situation and communicative needs. The separate language system that lies in every bi/multilingual minds are special and unique in their own way. There is no benchmark and it can never be compared to the 'ideal type' of language (Larsen-Freeman, 2011). The way language users use the language to suit their communicative needs and their experience has created a rhythm in the pattern of use of the language. The rise and the fall of the pattern of use of the target language is the proof that language learning process is dynamic and complex (Larsen-Freeman, 2006). Yet, this is not the whole story. We should keep in mind that some fresh findings and new interpretations of the existing results would be fundamental and may point towards important research in the future.

REFERENCES

Adel, A. (2015). Variability in learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.). *The Cambridge Handbook of Learner Corpus Research* (pp. 401-421). Cambridge: Cambridge University Press.

Arjan, A., Abdullah, N. H., & Roslim, N. (2013). A Corpus-Based Study on English Prepositions of Place, in and on. *English Language Teaching. 6*(12), 167-174.

Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing, 26,* 28-41.

Biber, D., S. Conrad & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Butler, Y. G. (2002). *Second language learners' theories on the use of English articles: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system.* Cambridge: Cambridge University Press.

Chapelle, C. A. (2013), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.

Chau, M. H. (2012). Learner Corpora and second language acquisition. In K.Hyland, M. H. Chau & M. Handford (Eds.), *Corpus applications in Applied Lingustics* (pp. 191-207). London: Continuum.

Chau, M. H. (2015). *From language learners to dynamic meaning makers: A longitudinal investigation of Malaysian secondary school students' development of English from text and corpus perspectives* (Unpublished doctoral dissertation): University of Birmingham, UK.

Cobb. T., & Horst. M. (2015). Learner corpora and lexis. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of Learner Corpus Research*. (pp.185-206). Cambridge: Cambridge University Press.

Cobuild, C. (2014). *Collins COBUILD advanced learner's dictionary: The source of authentic English* (8th ed.). Glasgow: HarperCollins.

Cook, V. (1991). Evidence for multicompetence. *Language Learning,44*(4), 557-591.

Cook, V. (2008). Multi-competence: Black hole or wormhole for second language acquisition research. In Z. HongHan (Ed.), *Understanding second language process* (pp. 16-26). New York: Multilingual Matters.

Cook, V. (2013). Multicompetence. In C.A. Chapelle (Ed.), *The encyclopedia of Applied Linguistics*. (pp. 3768-3774). Oxford: Wiley-Blackwell.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, *5*(1-4), 161-170.

Darus, S & Subramaniam, K. (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.

Diffloth, G. (1976). Jah-Hut, an Austroasiatic Language of Malaysia. In N.D. Liem (Ed.), *South-east Asian Linguistic Studies* (pp. 73-118).

Dong, G. (2013). On the nature of interlanguage. *Research on Humanities and Social Sciences*. *3* (18)*, 42-44.

Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press.

Ene, E. (2008). Developmental stages in advanced SLA: A corpus-based analysis of academic writing by ESL graduate students. *ITL International Journal of Applied Linguistics*, *156*(1), 53-86.

Esquina, A., Araujo, B., & de la Piedra, M.T. (2014). Meaning making and translanguaging in a two-way dual-languaging on the US Mexico border. *Bilingual Research Journal: The Journal of the National Association for Bilingual Education.* *37*(2), 164-181.

Fen-chuan, L. C. (2001). The acquisition of English articles by Chinese learners. *Second Language Learning.* *20*(1), 43-78.

Garcia, O. (2014). TESOL Translanguaged in NYS: Alternative perspectives. *NYS TESOL Journal, 1*(1), 2-10.

Garcia, O., & Sylvan, C. E. (2011). Pedagogies and practices in multilingual classrooms: Singularities in pluralities. *The Modern Language Journal, 95*(3), 385-400. doi: 10.1111/j.1540-4781.2011.0128.x0026-7902/11/395-400.

Granger. S. (1994). The learner corpus: A revolution in applied linguistics. *English Today. 10*(3)*,* 25-39. doi : 10.1017/ S0266078400007665.

Granger, S. (2003). The International Corpus of Learner English : A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL QUARTERLY. 37*(3), 538-546.

Granger. S., Gilquin. G. & Meunier. F. (2015). *The Cambridge Handbook of Learner Corpus Research.* Cambridge: Cambridge University Press.

Grosjean, F. (1989). Neurologists, beware! The bilingual is not two multilinguals in one person. *Brain Language, 36*(1), 3-15.

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning. 26*(2), 321-351.

Huebner, T. (1983). *A longitudinal analysis of the acquisition of English.* USA: Karoma Publishers Ann Arbor

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English.* Amsterdam: John Benjamins Publishing B.V.

Hunston, S. (2010). *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Klein, W. & Perdue, C. (1997). The Basic Variety (or Couldn't natural language be much simpler?) *Second Language Research*, *13*(4), 301-347.

Knowles, G. & Mohd Don, Zuraidah. (2005). Introducing MACLE: The Malaysian

Corpus of Learner English. In: *The first national symposium on corpus linguistic: Selected papers*. Wang Longyin & He Anping, Guong Zhou: North East Normal University Press.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2)*,* 141-165.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistic, 27*(4)*,* 590-619.

Larsen-Freeman, D. (2012). The emancipation of the language learners. *Studies in second language learning and teaching,2*(3)*,* 297-309.

Lennerberg, E.H. (1967). *Biologial foundations of language*. New York: Wiley.

Ministry of Education. (2013). Malaysia Education Blueprint 2013-2025. Retrieved from https://www.moe.gov.my/dasar/1207-malaysia-education-blueprint-2013-2025/file

Lightbown, P. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In H. Seliger and M. Long (Eds.), *Classroom Oriented Research in Second Language Acquisition* (pp. 217-243). Rowley, MA: Newbury House.

Lighbown, P. M. (1983b). 'Acquiring English L2 in Quebec classrooms'. In S. Felix and H. Wode (Eds). *Language Development at the Crossroads* (pp 101-120). Tubingen: Gunter Narr.

Master, P. (1997). The English article system. Acquisition, function and pedagogy. *System, 25*(2)*,* 215-232.

Master, P. (2002). Information structure and English article pedagogy. *System, 30*(3) 331-348.

McEnery, T., & Gabrielatos, C. (2006). *English corpus linguistics*. In Aarts, B. & McMahon, A (Eds.), The Handbook of English Linguistics (pp. 33-71). Oxford: Blackwell.

Meyer, Charles F. (2002). *English corpus linguistics*. Cambridge: Cambridge University Press.

Miller, J. (2005). Most of ESL Students Have Trouble with the Articles. *International Education Journal, 5*(5)*,* 80-88.

Morales, A. (2011). The role of the L1 in the acquisition of English articles by Spanish-speaking children. In J. Herschensohn and D.Tanner (Eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition conference (GASLA 2011)* (pp. 83-89). Somerville: Cascadilla Proceedings Project.

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance,and the bi/multilingual turn. *Language Learning, 63*(1), 1-24.

Ozfidan, B. & Burlbaw, L. M. (2019). A Literature-Based Approach on Age Factors in Second Language Acquisition: Children, Adolescents, and Adults. *International Education Studies, 12*(10), 27-36.

Pavesi, M. (1984). The acquisition of relative clauses in a formal and in an informal setting: further evidence in support of the markedness hypothesis. In D. M. Singleton & D. G. Little (Eds.), *Language learning in formal and informal contexts* (pp. 151-164). Dublin: Irish Association for Applied Linguistics.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics, 10*(1)*,* 52-79.

Romaine, S. (2006). *The bilingual and multilingual community*. In T.K Bathia & W. C Ritchie (Eds.), *The handbook of bilingualism* (pp. 385-405). Carlton: Blackwell Publishing.

Samad, A. A., Hassan, F., Mukundan, J., Kamarudin, G., Syed Abd Rahman, S. Z., Md.Rashid, Juridah, & Vethamani, M. E. (2002). *The English of Malaysian School Students (EMAS) Corpus*. Serdang: Universiti Putra Malaysia Press.

Samad, A. A. (2004). Beyond concordance lines: Using concordances to investigate language development. *Internet Journal of e-Language Learning & Teaching. 1*(1), 43-51.

Sawalmeh, M. H. M (2013). Error analysis of written English essays: The case of students of the preparatory year program in Saudi Arabia. *English for Specific Purposes World, 14*(40), 1-17.

Shamsudin, S., & Mahady, N. R. A. (2010). *Corpus linguistics based error analysis of first year Universiti Teknologi Malaysia students' writing*. (PhD thesis). Universiti Teknologi Malaysia, Kuala Lumpur.

Selinker, L. (1972). Interlanguage. *IRAL- International Review of Applied Linguistics*, *10*(1-4), 209-232.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse.* London: Routledge.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language learning. In H. Brynes (Ed.), *Advanced language learning: The contributions of Halliday and Vygotsky* (pp. 95-108). London: Continuum.

Tarone, E. (2006). Interlanguage. *Elsevier, 4*(1), 747-752. Retrieved from https://www.academia.edu/4788301/Interlanguage_Tarone_not_me_PDF

Tarone, E., & Swierzbin, B. (2009). *Exploring learner language*. Oxford: Oxford University Press.

Tarone, E. & Swierzbin, B. (2009). *Exploring learner language*. Oxford: Oxford University Press.

Velasco, P. & Garcia, O. (2014). Translanguaging and the writing of bilingual learners. *The Bilingual Research Journal: The Journal of the National Association for Bilingual Education. 37*(1)*,* 6-23. doi: 10.1080/15235882.2014.893270.

Yule, G. (2012). *The study of language.* Cambridge: Cambridge University Press