Collocations in a Learner English Corpus: Analysis of Yoruba-speaking Nigerian English Learners' use of Collocations

Peter Obukadeta

A Thesis Submitted to Kingston University London

in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

School of Arts, Culture and Communication

Kingston University London

May 2019

Abstract

The aim of the study reported in this thesis is twofold: to build a learner corpus of Nigerian English, and to investigate the production and use of collocations by Nigerian English learners. Computer learner corpora have offered us a new tool for better analysis and understanding of learner language enabling us to either reinforce or challenge some of our most-deeply rooted ideas about learner language. While learner corpus research has grown rapidly within its relatively short existence, there is no learner corpus of Nigerian English. This study built a half a million words Nigerian Learner Corpus of English (NILECORP) representing four proficiency levels (A2, B1, B2 and C1). While various studies have shown that learners have difficulties producing collocations, there has been a dearth of studies of collocations within the context of World Englishes. This study investigates the production and use of collocations by Yoruba-speaking Nigerian English learners not based on the notions of norms and standards of the prestigious varieties of English but based on the sociolinguistic reality of language use in the Nigerian context. Using LOCNESS (a native English corpus), NILECORP and the Nigerian component of the Corpus of Global Web-based English (GloWbE), this study investigates the extent to which native speakers and L2 learners use collocations, and the relationship between frequency of and exposure to input in L2 learners' speech community and their production of collocations. It also investigates the relationship between proficiency and the production of collocations, and the nature and causes of the collocational errors produced by the learners.

The findings suggest the difference between the collocations produced by the learners and the native speakers does not lie in the quantity but in the linguistic complexity – structural and semantic properties of the collocations produced. It also suggests that frequency and exposure to input facilitate the productive knowledge of collocations, and that frequency trumps incongruency. It shows that the production of collocations increases in tandem with proficiency increase but the production of congruent collocations decreases as proficiency increases. The most proficient group which produced more acceptable collocations than the others also produced the highest numbers of unacceptable collocations with L1 negative transfer being the biggest source of collocational errors across the four groups.

Acknowledgement

I am grateful to God for seeing me through all the difficulties I had throughout this PhD programme. I owe several debts of gratitude to my family, friends, colleagues and supervisors for their support. I want to express my sincere gratitude to my supervisors, Drs. Paul Booth and Clarissa Wilks, whose precious guidance, personal support, and inspiring feedback at all stages of the research design, data collection and the writing of this thesis have been invaluable spur and resources. I am extremely grateful to the Centre for English Corpus Linguistics, Université Catholique de Louvain, Belgium for giving me access to The Louvain Corpus of Native English Essays (LOCNESS) which is the primary reference corpus for this study. My gratitude also goes to the Students Life Centre, Kingston University for a Student Support Funds of £2,300.

My gratitude is due to my church community both in Nigeria and in the UK for their financial and moral support. First and foremost, I want to thank Mr and Dr (Mrs) Adeolu and Mojisola Adeniyi for their constant encouragement, prayers and financial support. They paid for my flight to the UK (in addition to other financial support) when coming for my Master's degree programme. I want to acknowledge the prayers and support of Pastor Adegboyega Adetoye who sadly died before the completion of this PhD. It is worth mentioning the prayers and constant encouragements of Adewumi Omikunle. I also owe a gratitude to my UK Pastor Olalekan Akinleye for his encouragements, prayers and financial support. Many thanks to Andy Okoro and Femi Sholagbade for their encouragement and financial support. I enjoyed the support of my church community throughout the PhD period, and for this, I am immensely grateful.

Finally, I am extremely grateful to my family. My beloved wife, Funmilayo, encouraged and prayed for me throughout this period. She endured all the hardship we faced in course of doing this PhD. I also thank my son, 'Little Emmanuel'. I am grateful to my brother, John Obukadeta, and my parents-in-law, Beatrice and Titus Olonipile for their supports.

Contents

Chapter One

Introduction

1.0 Introduction

There is a pronounced gap between L2 learners' collocational knowledge and their general linguistic knowledge (Bahns & Eldaw, 1993; Barfield, 2009). Unlike L2 speakers, native speakers of English are intuitively aware that some words in their language in some unspecified way tend to co-occur in a relatively fixed and recurrent combination, and by the same intuition would reject any violation of such lexical combination even when the resulting utterance seems to be grammatically correct and intelligible (Bartsch, 2004). Because the rules governing the co-occurrence of lexical items in collocations defy explanation based on regular combinational rules of syntax and semantics, collocations have become an inherent problem in second language acquisition.

Collocations, words that habitually appear together and thereby convey meaning by association (e.g. *deep remorse, strong evidence, adjourn meeting*), seemed to be a neglected variable in Applied Linguistics until Firth (1957) brought the concept into the limelight. Though a very important component of fluent linguistic production and a key factor in successful language learning, there is sufficient evidence in the literature that collocational deficiency is a pervasive linguistic phenomenon in second language acquisition (Bahns & Eldaw, 1993; Nesselhauf, 2005; Laufer & Waldman, 2011; Henriksen, 2013). Collocation has been a major area in vocabulary research which has attracted much interest since the late 1950s. Benson, Benson & Ilson (1997: ix) highlight the importance of collocation in second language acquisition by stating that if "learners wish to acquire active mastery of English, that is, if they want to be able to express themselves fluently and accurately in speech and writing, they must learn to cope with the combination of words into phrases, sentences and texts". This view lends credence to Lewis' (1993) Lexical Approach. The basic concept of the approach is based on the idea that an important part of language learning consists of being able to understand and produce lexical phrases as chunks.

Research on L2 collocational competence and production has increased tremendously in the field of Applied and Corpus linguistic as well as Natural Language Processing (NLP) from the 1990s to date. By Natural language processing, I mean the field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human languages. The increase in L2 collocation research is largely due to the availability of corpora and the increasing awareness of the significance of collocations in language acquisition. However, most of the studies focus on a few collocations or specific collocational types (Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Siyanova & Schmitt, 2008). Many of them are rather descriptive and lacking developmental focus, focusing more on the product of learning and not the actual process of acquisition. A good number of them are cross-sectional and exploratory, and very few are longitudinal studies (Li & Schmitt, 2010). Many of the researchers use various elicitation procedures with tasks types and testing instruments which make comparison across studies with same research aims difficult (Nesselhauf, 2005; Groom, 2009). This is partly due to the lack of standardised tools for measuring collocational knowledge. Very few studies have been carried out on developing standardised tools for measuring collocational knowledge (see Gyllstad, 2007: COLLEX and COLLMATCH; Revier, 2009: CONTRIX; Eyckmans, 2009: DISCO). The variation in the research methods and procedures, and the sometimes conflicting findings in collocational research is primarily a reflection of the lack of clarity and agreement on the theoretical assumption regarding the conceptualisation of the notion of collocations (Henriksen, 2013).

While the frequency-based collocational research tends to ignore the semantic analysis of the combinations, it focuses on performance rather than competence. The more subjective phraseological approach, on the other hand, focuses on the identification of combinations with clear semantic relations between the collocating structures. By so doing, it ignores the actual frequency of the use of collocations. All this is further complicated by various corpus-based studies which either fail to or vaguely define their corpus texts. This makes their findings almost meaningless in developmental terms, their findings cannot be reliably compared to findings on learners elsewhere, and their studies are difficult if not impossible to replicate in another context.

However, despite the increase in collocational research, there is paucity of studies on L2 learners' collocations competence and development of speakers of English as a second

language from Kachru's (1992) outer circle of World Englishes nations in Africa and Asia. Not much is known yet about the collocational competence and development of the learners and speakers of English as a second language in West Africa, and Nigeria in particular. Given the pace of technology (including corpus analysis software), the available of large corpora, and the relative ease with which we can now build a corpus, there remains plenty of scope for further work relating to L2 collocation research (Barfield & Gyllstad, 2009). So, this study is set against the backdrop of a dearth of study of collocations from the perspective of World Englishes – the emerging Englishes, in this case, Nigerian English.

It should be noted that the English we have in Nigeria is our second language; hence the participants in this study are learners of English as a second language. English is the language we use in our schools; we use it in the media and in our workplaces. We even use it for such cultural events as coronation of our traditional rulers. We use the language for everything, sometimes in conjunction with our various local languages. This is contrary to Quirk's (1985:6) position "that non-native speakers of English use the language in a narrow range of purpose". The Nigerian government's language policy and planning has been that every student be taught and able to use English as an effective means of communication in any given situation. To this end, the four groups of participants in this study – Yoruba-speaking secondary school student learners of English as a second language, between the ages of thirteen and seventeen are expected, upon completion of their secondary school education, to be able to communicate effectively in English in any given context. They are expected to have the ability to express themselves clearly and coherently in a manner that is appropriate to the audience, purpose, topic and situation. It is expected that they should be able to use the language in a way that reflects the cultural specifics of language use appropriate to the Nigerian communicative context. They are expected to have such proficiency that is sufficient to communicate with any English-speaking person around the world. All the participants can achieve the above to varying degrees in proportion to their proficiency level. However, a substantial number of secondary school leavers do not pass English in their secondary school certificate examination and hence cannot to proceed to the university and other higher education institutions because a credit pass in English is required. Collocational deficiency could be a factor in the students' performance in their final English language examination. Various researchers (Benson, 1985; Brown, 1974; Cowie, 1981; Lewis, 1997) have highlighted the importance and value of collocations for the development

of L2 vocabulary and communicative competence. Brown (1974), in particular, points out that collocations enhance improvement of learners' oral fluency, listening comprehension as well as reading speed. I will now discuss the aims of the research reported in this thesis.

1.1 The aims of the Thesis

The aim of this study is twofold: (a) to build a learner corpus of Nigerian English – the first of its kind and (b) to investigate the production and use of collocations by Nigerian English learners. While various studies have shown that learners have difficulties producing collocations, we have not really explored the difficulties English speakers from the context of World Englishes have. Most especially, the difficulties speakers of English from the former British colonies where we now have new varieties of English face while producing collocations. Neither have we explored these new Englishes to see if there are certain collocations which are peculiar to these varieties of Englishes. And we do not have a learner corpus of Nigerian English with a clear definition of the proficiency levels the corpus texts represent. This study, therefore, intends to initially build a half a million words learner corpus and later expand it to a ten million words learner corpus of Nigerian English with clearly demarcated different proficiency levels and different Nigerian L1s so that researchers can compare learners from different Nigerian L1s at various proficiency levels.

The second aim of this study which is to investigate Nigerian learners' collocational knowledge and development is fourfold. (1) To investigate and compare from various perspectives the extent to which native and non-native writers make use of collocations in a written text (2) To explore, from various perspectives, the impact of frequency of and exposure to input in the learners' speech community on the production of collocations. Most L2 collocational studies have investigated the effect of frequency on the production of collocations within instructed language learning setting but this study attempts to investigate the effect of frequency of and exposure to input outside the classroom. (3) To investigate, from various perspectives, the relationship between proficiency and the production of collocations across various proficiency levels. (4) To analyse all the unacceptable

collocations produced by the learners; to identify, classify and account for the errors using appropriate language acquisition models. Basically, this study aims to investigate the collocational competence and development of speakers of English as a second language as opposed to English as a foreign language. By English as a second language, I mean in a context where a new variety of Standard English (not Pidgin English) is both the official language and lingua franca as in the case of former British colonies like Nigeria, Ghana, etc.

Meanwhile, L2 collocation studies in the literature have been based majorly on two conceptual underpinnings: frequency-based and the phraseological traditions (Barfield and Gyllstad, 2009). In frequency-based studies, frequency and statistics are intrinsic ingredients in the analysis of textual instantiations of collocations while research on collocation based on the phraseological tradition is guided by syntactic and semantic analysis. Collocations in the frequency-based tradition are viewed as units consisting of co-occurring words within a certain distance of each other (Firth, 1961). So, from the perspective of this conceptual underpinning, collocation is essentially a matter of frequency of co-occurrence, but this is not the case with the phraseological tradition. Contrary to the frequency-based approach, studies within the phraseological approach are based on the treatment of collocation as word combination, displaying varying degree of fixedness and in the preoccupation with the decontextualized classification of collocation. While frequency of co-occurrence matters much in collocation research, focusing on frequency alone may be inadequate in researching the complexity of L2 collocations. On the other hand, the phraseological tradition of treating collocation based on the degree of the fixedness of the co-occurring words while ignoring the frequency of co-occurrence appears inadequate as well. In view of the foregoing, this study seeks to investigate L2 collocation using a hybrid method – a combination of frequency-based and phraseological approach. I will provide my definition of collocation after reviewing the existing literature on the general phenomenon of collocation and studies on L2 collocations.

Before embarking on this thesis, I conducted a pilot study to explore the productive collocational knowledge of two groups of Nigerian advanced speakers of English as second language. The population of the study consisted of sixty educated Nigerians: thirty of them had been living in the UK for up to twenty years (some of them had done their postgraduate

studies in the UK) while the other thirty had never lived or studied outside Nigeria. I got my motivation for this pilot study from the growing body of evidence in the literature which suggests L2 collocation is a problematic linguistic phenomenon (Bahns & Eldaw, 1993; Farghal & Hussein, 1995; Nesselhauf, 2004, 2005; Siyanova & Schmitt, 2008; Wolter & Gyllstad, 2011). Of particular interest to me were two corpus-based studies conducted in Germany and Sweden by Nesselhauf (2005) and Groom (2009) respectively. While Nesselhauf used the German Corpus of Learner English (GeCLE), a precursor of the German component of the International Corpus of Learner English, Groom used Uppsala Student English Corpus (USE) – a 1.2m words corpus of undergraduate student essays written by Swedish university students compiled by staff of Department of English, Uppsala University, Sweden (Groom, 2009). Nesselhauf (2005: 236) who investigates the use of collocations by German advanced learners of English reports that "the length of stays in English speaking country does not seem to lead to an increased use of collocations; instead, there even seems to be a slight trend in the opposite direction". This seems to suggest that collocation is such a problematic linguistic phenomenon that even living in the target language context where the learner is supposed to have maximum exposure to the target structures may not necessarily translate to accelerated acquisition. More importantly, Nesselhauf (2005) has called into question the traditional belief that the best way to develop a native-like command of a second language (collocations) is to spend an extended time in the target language environment.

But Groom (2009: 30) who investigates the effect of second language immersion on L2 collocational development reports that "collocational usage and time spent in the target language context are more positively than negatively correlated". This, apparently, negates Nesselhauf's findings but it is very important to note at this stage that these two researchers belong to two different schools of thought on the theoretical notion of collocations. While Nesselhauf views collocations from the phraseological perspective, Groom is firmly rooted in frequency-based approach. This would obviously have influenced both the methods and the analytical framework they have adopted in their study which could explain why they came up with two opposing conclusions. Though they disagree on the correlation between second language immersion and L2 collocational usage, Groom (2009: 33) in his overall conclusion acknowledges that "the process of L2 collocational development is likely to be a slow and occasionally painful one quite irrespective of the linguistic environment in which the learner happens to be immersed". So, it could plausibly be concluded that irrespective of one's

theoretical perspective of collocations, L2 collocational deficiency is a pervasive phenomenon in second language acquisition, and immersion is not necessarily a solution to the problem. All this left me with many unanswered questions about L2 collocational competence and development, particularly, the collocational production and processing of speakers and learners of English from Kachru's (1992) outer circle of World Englishes.

In view of the above, the pilot study which is a prelude to this thesis was aimed at exploring the collocational competence of Nigerian advanced speakers of English as a second language – a context where English is the principal lingua franca of educated Nigerians, the principal medium of instruction in schools, the principal medium of wider communication, and the principal medium of literary expression. So, I wanted to find out if Nigerian Advanced Speakers of English would have problem producing collocations; and if so, I wanted to inquire into which types of collocations were more problematic for them. In addition to these, I also wanted to know the effect of long stay in the UK (English as a native language context) on their collocational competence. This I did by comparing the collocational competence of the two groups. My findings, which I will discuss in detail in chapter four, suggest that to a significant extent, collocation is a source of difficulty for Nigerian advanced speakers of English particularly incongruent collocations. Collocations are categorized as congruent and incongruent collocations based on the presence or absence of a literal L1 translation equivalent. Collocations that have lexical components that are similar in L1 and L2 are congruent collocations while the ones that have lexical components that are different in the two languages are incongruent (Nesselhauf, 2003; Yamashita & Jiang, 2010).

Going by the findings of the pilot study, one might conclude that the major cause of collocational deficiency is L1 transfer. However, a study by Wang and Shaw (2008) reveals that two groups of participants – one with Chinese as L1 while the other had Swedish as L1 made similar types and proportions of errors despite having different L1 and obviously having different incongruent collocations. This suggests that intralingual factors are as important as L1 factors when considering the potential sources of collocational errors. So, when most of the participants in my pilot study had problem producing incongruent collocations, I thought there might be more to it than L1 transfer – maybe some yet to be

identified factors are responsible for this performance. But, of course, all this leaves many questions unanswered.

Another discovery which I made, a very startling one, was that the participants who had never lived or studied outside Nigeria produced more acceptable collocations than the other group of participants who had been living in the UK – the target language environment – for up to 20 years. This finding, which is counterintuitive, throws up many questions which need to be empirically investigated. Why would speakers who are living in the UK, who are supposedly exposed more to the so-called native English produce fewer acceptable collocations? On the contrary, they produced more unacceptable collocations. Going by this finding which seems to corroborate an earlier finding by Nesselhauf (2005), one would seriously question the traditional assumption that the best way to develop a native-like command of a second language is to live and/or study in the target language context. I became more curious considering the fact that there is a gap in the literature regarding studies from the outer circle of World Englishes particularly in Nigeria where only a handful of studies have been carried out (Taiwo, 2001, 2004, 2010; Akande, Adedeji & Okanlawo, 2006; Israel, 2014) and none of them is corpus-based. Above all, I wanted to use a method that rules out some of the intervening variables in collocational research so as to either reinforce or challenge some of the theoretical issues around L2 collocational acquisition.

Using the findings of the pilot study as a lunch pad, in relation to the aims of this study articulated earlier, this research is focusing on four broad questions based on the hypothesis that second language learners inherently have problem producing collocations. Since the pilot study suggests that Nigerian advanced speakers of English have difficulties producing collocations, it is only plausible to look at the situation with Nigerian learners. If I were to continue and expand the pilot study as part of my main research, I might not be able to have an understanding of the acquisition process Nigerian learners go through before they reach the advanced speakers' stage. Looking at both advanced speakers and learners at the same time might be too ambitious and unmanageable. Guarded by the findings of the pilot study, this study will endeavour to answer the following questions:

1. To what extent do native and non-native writers make use of collocations?

2. Is there a relationship between frequency of and exposure to input in L2 learners' speech community and their production of collocations?
3. What is the relationship between proficiency and the production of collocations?
4. What is the nature and causes of the errors in the collocations produced by the learners?

I used a corpus-based method to achieve the aims by first comparing the collocations in the learner corpus - the Nigerian Learner Corpus of English (NILECORP) and the Louvain Corpus of Native English Essays (LOCNESS) which is the main reference corpus. NILECORP, the half a million words learner corpus I built, is made up of four sub-corpora representing four different language proficiency levels while LOCNESS**,** on the order hand, is a corpus of native English essays. I also compared frequency data from NILECORP and the Nigerian component of the Corpus of Global Web-Based English (GloWbE) which is the secondary reference corpus. I will expand on this in the methodology chapter.

1.3 Map of the Thesis

The introductory chapter focused on a general introduction to the thesis. It dealt with the statement of the problem and highlighted the gap in the literature on L2 collocations research pointing out how collocations in World Englishes have been neglected. It stated the aim of the study and the research questions, providing a highlight of the pilot study which is precursor to the main study. The first chapter was concluded with a brief description of the method used in this study.

The second chapter contains a review of the existing literature. It starts with a review of the literature on the general phenomenon of collocation, tracing the establishment and development of the concept of collocation in linguistic theory. I reviewed the literature on the main theoretical frameworks within which the concept of collocation has been addressed so far in the linguistic literature. This review includes the numerous and sometimes conflicting definitions of collocation in the existing literature. The various defining criteria (qualitative, quantitative and positioning criteria) were also examined. There is also a review

of the linguistic descriptions of collocation that have been provided in the literature. The focus is on the semantic compositionality and morpho-syntactic characterisation of collocations as well as the literature on classification of collocations. The literature review will then moves on to reviewing the existing literature on L2 collocations research. The review is divided into two parts. The first part focuses on studies on L2 collocation competence and development elsewhere while the second part focuses on studies on collocations in Nigeria – both collocations in L2 English and L1 Yoruba which is the language of the participants in this study. The literature review concludes with a review of the existing literature on Learner Corpus Research (LCR) and Nigerian English. The review of the literature on LCR is limited to learner corpus design and development, methodological issues and applications, particularly its application to L2 collocations research. The review of the literature on Nigerian English highlights the features that distinguish it from other varieties of English which means the possibility of the existence of collocations peculiar to the Nigerian speech community which the existing literature of L2 collocations research have not accounted for.

The third chapter focuses on the pilot study. It details the background to the study, the aims and the research questions. The chapter spelt out the methods and procedures used in the study. It ends with a presentation of the results and a discussion of the findings as well as a description of how the pilot study helped to shape the design of the main study.

The fourth chapter is divided into four sections. The first section presents the four main research questions and their sub-questions. The second section provides an overview of the various methods that have been used in L2 collocation research, focusing on corpus-based method and providing the justification for using corpus-based approach in this study. The study corpus – the Nigerian Learner Corpus of English (NILECORP) – is also presented in the second section. It also contains the explanation of the design criteria and the procedures for building the study corpus from ethics approval to defining and describing the population as well as data elicitation, data capture and text handling. The mechanism for converting the hand-written texts into electronic format and the assignment of proficiency levels to the corpus texts. The third section of the chapter also focuses on the reference corpora – The Louvain Corpus of Native English Essays (LOCNESS) which is the primary reference

corpus, and the Nigerian component of the Corpus of Global Web-based English (Davies, 2013) – the secondary reference corpus. The third section ends with the justification for using these corpora as the reference corpora for this study. The last section of the fourth chapter describes the approaches and procedures used for the extraction of collocational candidates from the study corpus and the primary reference corpus as well as the analytical approaches used for analysing the data and how each aspect of the research method addresses my research questions.

The presentation and analysis of data start in chapter five. This chapter investigates and compares the extent to which native and non-native writers make use of collocations in a written text considering four questions. It provides a detailed comparative analysis of all the Verb Noun and Adjective Noun collocations produced in the LOCNESS and NILECORP-C1. NILECORP-C1 is the most proficient of the four sub-corpora used in this study. This learner group is equivalent to the Common European Framework of Reference for Languages proficiency level C1. This chapter is divided into five sections. The overall descriptive statistics of the data used for the first main research questions and its sub-questions are presented in section one. The second section focuses on the comparative analysis of the linguistic complexity of the verb noun collocations produced by the native speakers and the L2 learners in terms of the collocation span and the structural properties of their constituents. The third section on the other hand focuses on the extraction and analysis of collocations which have had their meanings modified to introduce additional nuances and associations with the aim of understanding the extent to which L2 learners produce and use semantically opaque collocations with varying degree of idiomaticity. The congruent and incongruent collocations produced by the learners are analysed in the fourth section. This chapter ends with a discussion section focusing on interpreting and explaining my findings and examining whether and how my research questions have been answered. The discussion shows how my findings relate to the immediate literature on native speakers and L2 learners' use of collocations.

Chapter six further elaborates the analyses reported in chapter five by considering how frequency and exposure to input in the learners' speech community affect the collocational production of L1 Yoruba learners of English. Frequency data from the Nigerian component of GloWbE was used to determine the effect of frequency of the collocations produced. The

chapter also ends with a discussion section showing how my findings relate to the immediate literature on the effect of frequency on the acquisition of collocations.

Chapter seven enquires into the relationship between language proficiency and the production of verb noun and adjective noun collocations across four groups of L2 learners representing four different proficiency levels. There is a focus on the relationship between proficiency and the use of linguistically complex verb noun collocations in terms of the collocation span and the structural properties of their constituents. It also addresses the relationship between proficiency and the use collocations with additional nuances and associations – the degree of semantic opacity and transparency. The aim is to find out if L2 learners' knowledge of collocations increases in tandem with their general proficiency in the English language. This chapter also ends with a discussion showing how the findings relate to the immediate literature on the relationship between proficiency and L2 learners' use of collocations

Chapter eight inquires into the collocational errors produced by the L2 learners. The notion of errors in this study is not based on the notion of norms and standards of the prestigious varieties of English but, on the contrary, based on the sociolinguistic reality of the English language use in the Nigerian context. The focus of the error analysis is on the identification, classification and the analysis of all the erroneous verb noun and adjective noun collocations extracted from the four sub-corpora. It addresses four broad questions related to the errors extracted from the corpus texts. This chapter also ends with a discussion on the errors within the literature on collocational errors.

All the themes that emerge from the study are discussed in chapter nine. The chapter is organised into two parts. The first part focuses on the learner corpus, collocations in World Englishes and the question of norms and standards in the English language with specific focus on collocations in Nigerian English. The second part of the chapter discusses the collocational errors further considering the role of interlexical and intralexical factors in the production of collocations focusing on clang associations, frequency of input and congruency; and attempts to explain collocational links in L2 mental lexicon. The chapter

ends with a discussion of the findings within Usage-based theory of language acquisition (Tomasello, 2003) and Jiang's (2000) Model of Vocabulary Acquisition.

The tenth chapter concludes the study by presenting the summary of the findings and showing the extent to which the aims of this study were achieved. It discusses how significant the results are as well as the limitations of this research. It points out some interesting further areas to be explored based on the findings. Finally, it provides some recommendation on the teaching of collocations in Nigeria.

Chapter Two

Literature Review

2.0 Introduction

This chapter's main purpose is fourfold: (a) to review the literature on the English language in Nigeria (b) to review the literature on the general phenomenon of collocation (c) to review the existing literature on L2 collocation research and (d) to review the literature on Learner Corpus Research (LCR) that is relevant to this study. This literature review starts with the existing literature on Nigerian English highlighting the features that distinguish it from other varieties of English. Since this study investigates the collocational knowledge of learners of English from a World Englishes perspective, it is important to shed light on the features of Nigerian English as a variety marker. This is necessary to help understand the context and make sense of the findings of this study as the decision on whether the collocations produced by the participants are acceptable or unacceptable is not based on the norms and standards of the prestigious varieties of English but on Nigerian English. The issue of norms and standards will be discussed later in the thesis.

Before reviewing the literature on L2 collocation research, which is the second purpose of this chapter, it seems appropriate to review the literature on the general phenomenon of collocation first to provide the context for the review of the literature on L2 collocation research. The concept of collocation is not too popular in mainstream linguistics, nor is it too well understood. It is still somewhat vague despite the increase in collocation research mainly due to the availability of corpora and corpus analysis tools. Yet, as Benson et al (1986a: vii) put it, knowing your collocation is "of vital importance to those learners of English who are speakers of other language". Collocation is so important for fluent linguistic production that same holds for native speakers:

> "In order to speak natural English, you need to be familiar with collocations. You need to know, for example, that you say 'a heavy smoker' because heavy (NOT big) collocates with smoker, and that you say 'free of charge' because free collocates with charge (NOT cost, payment, etc.). If you do not choose the right collocation, you will probably be understood but you will not sound natural" (Longman Dictionary of Contemporary English, 1987:193).

In spite of this fact, collocation is rarely treated in the theoretical literature. In view of the foregoing, I will discuss collocation from as many viewpoints as possible, taking into consideration a substantial amount of literature. I will start by tracing the establishment and development of the concept of collocation in linguistic theory. This will be followed by a review of the main theoretical frameworks within which the concept of collocation has been addressed so far in the linguistic literature. These theoretical perspectives will include: Contextualism, Text Cohesion, Meaning-Text Theory (Firth, 1957; Mel'čuk, 1981; Halliday & Hassan, 1976; Žolkovskij & Mel'čuk, 1967; Seretan, 2011). Having done this, I will then review the numerous and sometimes conflicting definitions of collocation in the existing literature. The various defining criteria (qualitative, quantitative and positioning criteria) will be critically examined. This section will also include a review of the linguistic descriptions of collocation that have been provided in the literature. The focus will be on the semantic compositionality and morpho-syntactic characterisation of collocation. Because of the nature of the research which is being reported in this thesis, it is important to review the literature on classification of collocations. These classifications will include: BBI Classification, Mel'čuk's Classification, Aisenstadt's Classification, Hausmann's Classification, and Cowie's Classification (Benson et al, 1986a; Nesselhauf, 2005). After this extensive review of the literature on the phenomenon of collocation, I will then focus on reviewing the existing literature on L2 collocation research.

After the review of the literature on the general phenomenon of collocation, the focus will then be on studies on L2 collocation knowledge and development which will be divided into two main sections. The first section focuses on L2 collocations across the world while the second section will focus on studies on collocations in Nigeria – both collocations in L2 English and L1 Yoruba. This literature review chapter will be concluded with a review of the literature on Learner Corpus Research (LCR). LCR is a relatively young but vibrant new branch of research. It stands at a crossroads between corpus linguistics, foreign language teaching and second language acquisition (Granger, Gilquin & Meunier, 2013). Its origins could be traced back to "the late 1980s when academics and publishers, concurrently but independently, started collecting data from second language learners with the purpose of advancing our understanding of the mechanisms of second language acquisition and/or developing pedagogical tools and methods that more accurately target the needs of language learners" (LRC Conference, 2011) rather than just depending on intuition. Considering the

scope of this study, the review will be limited to the main aspects of learner corpus research that are relevant to this study.

2.1 English Language in Nigeria

The world has previously witnessed the spread of languages of empires (e.g. Latin, Greek, Aramaic, etc.), the diffusion of lingua franca and the growth of international languages (Fishman, 1992). But it is nothing compared to the continuous spread of the English language for international and intranational purposes. Within the last century, the English language has changed the linguistic ecology of the world; and no country, obviously, feels the impact of this linguistic revolution more than the former British colonies of which Nigeria is one. The English language first came in contact with the people of the southern coast of what is now modern-day Nigeria around 1553 (Spence, 1971) initially through English traders, then explorers, anti-slavery activists, missionaries, and finally entrenched through British colonial rule. In less than a hundred years, the English language has altered the Nigerian linguistic landscape. Today, we have a new sociolinguistic reality – the emergence of a new variety of the English language. Over the years, the English language having come in contact with new cultures and literature has evolved to accommodate lexico-semantic, discourse, phonological and grammatical features that are in tandem with the sociolinguistic reality of language use in Nigeria. This new variety of English, with its characteristic accents, syntactic features, lexis, pragmatic features and the like reflects the people's local linguistic and cultural influence on the English language that was brought to us by the colonial masters. We now have a "new English, still in communion with its ancestral home but altered to suit its new [Nigerian] surroundings" (Achebe, 1976:11). This new variety of English is now widely referred to as Nigerian English. There is more than one variety of English in Nigeria, but the focus of this study is on the Nigerian Standard English. This is the variety that is used in educational and official settings. The other variety – the Nigerian Pidgin English, though widely used across the country, is, however, not used in official setting. But is there indeed any such thing as "Nigerian English"?

Much has been written about the existence of Nigerian English since the time when Walsh (1967 in Ogu, 1992:88 cited in Ajani, 2007) drew attention to the fact that: "the varieties of

English spoken by educated Nigerians, no matter what their language, have enough features in common to mark off a general type, which may be called Nigerian English". However, there is no unanimity in the assessment and definition of Nigerian English (Bamigbose, 1982; Tijani, 2007; Kporegi, 2007; Christiana-Oluremi, 2013; Okurinmeta, 2014) hence no one has been able to come up with a universally acceptable definition. This is probably due to the complex and evolving nature of Nigerian English engendered by the continuous influence of the various local languages. There does not seem to be a single definition that encompasses the entire spectrum of Nigerian English. Kperogi (2007) defines Nigerian English as the variety of English that is broadly spoken and written by Nigeria's literary, intellectual, political, and media elite across the regional and ethnic spectra of Nigeria. But a cursory look at this definition will quickly reveal it is problematic. It does not tell us how Nigerian English is different from the other Englishes and what qualifies it as 'Nigerian'. Odumah (1987 cited in Ajani, 2007) simply identifies Nigerian English vaguely as one of the new varieties of the English language developing around the World. He proceeds to sub-divide Nigerian English into three dialects arising from the influences of the three major Nigerian languages – Yoruba, Hausa, and Igbo. While this is true to some extent, this categorization alienates other local varieties used in other speech communities where Yoruba, Hausa, and Igbo are not L1. Bamigbose (1982: 105 cited in Ajani, 2007), a respected Nigerian linguist, views Nigerian English as the English which local educated Nigerians use in "natural and spontaneous usage".

While I recognize the fact that the English language as used by educated Nigerians could be used as the benchmark for what counts as Standard Nigerian English, this is somehow problematic in the sense that there is varying degree of education. When Bamigbose says 'educated Nigerians' what level of education is he referring to? Is he referring to university level education or secondary school level as both of them may be referred to as educated? If we say university educated, what if such an 'educated Nigerian' uses the English language in a way that appears to deviate from the 'accepted norm' of what we now refer to as Nigerian English, do we count it as error or innovation? And where do we draw the line between usages that are genuinely Nigerian in nature and those that are outright errors of usage?

A much more comprehensive definition of Nigerian English which I will use as a working definition in this study is the one advanced by Osunbade (cited in Christiana-Oluremi, 2013: 264). He asserts that "Nigerian English is, therefore, that variety of English that has

developed in the Nigerian non-native situation and it has distinguishing features manifested at the phonological, lexico-semantic, grammatical, and discourse levels". So there exists indeed a Nigerian English which is identifiable. Despite the local varieties, there is at the moment a single super ordinate variety of Standard English in Nigeria which can be regarded as Nigerian English (Odumah, 1993). This English is indigenous to Nigeria and its basic usage is intra-national (Ajani, 2007). It is mainly distinguishable from other Englishes through its semantic component.

The task of identifying, isolating and accounting for the linguistic features of Nigerian English as variety markers has been described as elusive (Kaan, Amase & Tsavmbu, 2013). This is more so because as Kaan, Amase & Tsavmbu (2013: 76) observe, "the English language in Nigeria has been cultivated and re-domesticated as well as indigenized to accommodate the culture and tradition of the people and as such, has acquired local colour and distinguished itself from the native speaker variety with features reflected at the semantic level". But the distinguishing features of Nigerian English are not limited to the semantic level. While the English language has been influenced at every level – syntactic, pragmatic, lexical, phonological and semantic – by the Nigerian socio-cultural environment, the semantic level of the language seems to be the most susceptible to creativity in Nigerian English language usage context (Kaan, Amase & Tsavmbu, 2013). They note that "semantic variation has been a pervasive characteristic of the Nigerian variety of English" (Kaan, Amase & Tsavmbu, 2013: 80). Considering the scope of this study, the literature review is limited to the semantic and syntactic features of Nigerian English. The semantic and syntactic features are likely to reflect the peculiarities of the collocations in Nigerian English.

Identification and description of what constitutes Nigerian English has been the subject of many studies (Odumuh, 1983; Jowitt, 1991; Bamigbose, 1995; Bamgbose, Banjo & Thomas, 1995; Ajani, 2007; Kaan, Amase & Tsavmbu, 2013; Anyachonkeya & Anyachonkeya, 2015). There is unanimity in the literature on the semantic features of Nigerian English that it is characterised by "meaning narrowing, semantic extension, semantic reduplication, semantic shift, coinage of new words with new meanings, the Nigerianisation of idioms and proverbs, ambiguity resulting from omission of articles among other semantic issues as marking off Nigerian English" (Kaan, Amase & Tsavmbu, 2013: 76). Studies on the syntactic features of

Nigerian English equally show features that mark it out as a different variety. A Survey of the Syntactic Features of Educated Nigerian English by Edem (2016: 1) reveals a "very slight variation in Nigerian English at sentence, clause, group and word levels from the structures of the British English usage". He concludes there is not much difference between the syntactic structures of educated Nigerian English and British English. Although it is not clear how he compared the syntactic features of Nigerian English with British English to be able to reach such conclusion, his findings are however, consistent with Jowitt's (1991: 109) earlier findings that "the gap between Educated Nigerian English syntax and Standard British English syntax when each is considered in its entirety is narrow, not wide."

The various studies that have been reviewed clearly point to the existence of a variety of English that is distinct from the other varieties of the language. But despite the nativization of English in Nigeria – now possessing the colouring of the immediate speech community, it still shares common core features (like common grammatical rules) with the prestigious varieties of English. Finally, on this section, as Bamgbose (1995) rightly points out, the nativization of English in Nigeria is not limited to the features of L1 transfer. On the contrary, the nativization also involves the creative use of the language as well as the evolution of the unique pragmatic usage of the language in a way that reflects the sociolinguistic reality of language use in Nigeria. This transformation may have resulted in the development of certain collocations that are peculiar to the Nigerian context which the existing literature on collocations based on the norms and standards of the prestigious varieties of English have not accounted for.

2.2 The Establishment and Development of the Concept of Collocation

As a port of departure, it would be helpful when beginning a section on such an important concept to this thesis, to provide a simple and unambiguous initial definition of collocation. After a thorough review of how collocation is different from other recurrent word combinations and how collocation has been construed in the various literature on collocation, a final definition of collocation for the purpose of this thesis will be presented at the end of

this chapter. Even a cursory glance at the literature on the concept of collocation will reveal that forming a definition that will be precise enough and yet capture all the core elements of collocation is difficult. Hence the literature is rife with both conflicting definitions and conflicting terminologies. This is attributable to the fact that "collocation is a term which is used and understood in many different ways" (Bahns, 1993:57). In essence, the concept of collocation is somewhat vague but despite the variations, a workable definition can be formed. In doing so, some key factors regarding the concept of collocation that are central to later analysis will be considered. These factors will include the development of the concept of collocation in linguistic theory, and some key elements of collocation such as the notion of collocational span.

While collocation has been variously defined as a lexical, grammatical or research phenomenon (Brown, 1974; Kjellmer, 1987; Scott, 1999), all the definitions, as varied as they are, focus on the co-occurrence of words. Firth (1957:179) in his study of collocation, declares that "you shall know a word by the company it keeps". This is obviously a reference to words that habitually appear in the company of certain words. In the same vein, Kjellmer (1987:133) defines collocation as "a sequence of words that occurs more than once in identical form and well structured". This is similar to Clear's (1993:277) view of collocation as "a recurrent co-occurrence of words". The common core of agreement in all the above definitions is the focus, implicitly or explicitly, on recurrent co-occurrence of words. It should be noted that it is not every group of words that habitually co-occurs and apparently belongs to set of ready-to-hand units of language that are collocations. But at this stage of the thesis, I will initially define collocation as words that keep company with one another.

Collocation is a complex concept. It is, therefore, important to discuss the establishment and development of collocation in linguistic theory before going into detailed discussion on its key elements. The term collocation has been used in linguistic context since 1750 (Bartsch, 2004). In the second edition of Oxford English Dictionary, a quotation by Harris made a reference to it as follows: "the accusative …in modern languages … being subsequent to its verb, in the collocation of the words" (Harris, 1750 cited in Bartsch, 2004:28). In the above quotation, the term is used in a sense that is quite different from how it is used now. There is nothing in the quotation that suggests the strongly lexical character now associated with the

concept of collocation over and above the grammatical relation between the constituent parts. Harris used the term in a sense that is now widely covered by the closely related term *colligation.* Colligation is the grammatical juxtaposition of words in a sentence (Bartsch, 2004). It denotes the grammatical relation between lexical items. Collocation on the other hand, as it is currently used entails a grammatical relation between lexical items as well as particular co-selection constraint on the choice of lexical item that can co-occur (Pawley & Syder, 1983). In another quotation cited in the same dictionary, Trager in 1940 used the term collocation to denote the general combinatorial properties of linguistic elements – not limited to lexical items (Trager, 1940).

In the 1930s, Palmer (1933), who is widely regarded as the pioneer of the field of English as a Foreign Language recognised the importance of collocations in language learning and the need to teach them. He built a list of 6,000 frequent collocations (Seretan, 2011). This is obviously a very significant contribution to the study of collocations. He was perhaps the first to pay attention to collocations and includes them in his teaching materials and thought they be taught as one linguistic element. However, the contribution of Palmer to collocational studies is often overlooked and overshadowed by the contribution of Firth. Over time, it was becoming obvious that the phenomenon of collocation was vital component of language. In recognition of this fact, A. S. Hornby included collocational information in the dictionaries from the series he initiated. The dictionaries with collocational information include: Idiomatic and Syntactic English Dictionary (Hornby, 1942), Oxford Advanced Learner Dictionary (Hornby et al, 1948a), and The Advanced Learner Dictionary of Current English (Hornby et al, 1963).

The literature widely credited Firth with systematically introducing the concept of collocation into linguistic theory. He was among the first linguists to base a theory of meaning on the notion of "meaning by collocation" (Firth, 1957). He proposed to bring forward as a technical term, meaning by 'collocation' and to apply the test 'collocability' (Firth, 1951; 1957). He explains the term collocation in more details:

> "Meaning by collocation is an abstraction on the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of night is its collocability with dark, and dark, of course, collocates with night" (Firth, 1951 cited in Schiebert, 2009: 3).

He was largely responsible for channelling the attention of linguists towards lexis and actually popularised the concept of collocation. As Krishnamurthy (2000) rightly points out, he is credited for establishing the distinction between cognitive and semantic approaches to word meaning on the one hand, and the linguistic features of collocation on the other hand. He was convinced that language should be studied as a social phenomenon by regarding its social context beyond the purely linguistic facts. In the light of this, collocation plays a central role in contextually determining meanings. Firth was not alone in this view, Palmer (1933), Porzig (1934 cited in Seretan, 2011), and Coseriu (1967 cited in Bartsch & Evert, 2014) also advocate the view that the meaning of a word is established by its co-occurrence with particular other words in the same context. This line of thought will be discussed further under Contextualism as one of the theoretical perspectives on collocational research. Meanwhile, Firth went on to point out that collocation has to be observed in connection with specific registers, genres, authors, and texts (Schiebert, 2009).

There seems to be some contradictions in the literature about who coined the word 'collocation' and who was actually the first linguist to use the term collocation in the sense of a recurrent, relatively fixed word combination. There have been claims in the literature that the word 'collocation' was coined by Firth (Schiebert, 2009). But contrary to such claims, Palmer (1938) in his book "A Grammar of English Words" used the term 'collocation'. While explaining what collocation is and how collocations are treated in his 'grammar of words' he stated that:

> "When a word forms an important element of a 'collocation' (a succession of two or more words that may best be learnt as if it were a single word) the collocation is shown in bold type [...]. The collocations are entered so far as possible under the appropriate semantic variety of the word […].
>
> When, however, the meaning of the word in the collocation (or group of collocations) differs considerably from any of the meaning listed under 1, 2, 3, etc., and independent paragraph is provided (Palmer [1938] 1968: x cited in Bartsch, 2004: 32).

The above quotation suggests that someone else might have coined the term collocation and not Firth. Evidence in the literature as discussed earlier indicates that the term has been in use in linguistic context before Firth brought it to the limelight. Palmer's reference to 'words that may best be learnt as if it were a single word' suggests that he used the term in the sense of recurrent, relatively fixed word combinations. But there is no evidence to conclude that he

was the first linguist to use the term collocation in this context. One thing that is apparently indisputable is the fact that Firth and his successors, the so-called Neo-Firthians played significant role in establishing the concept of collocation in linguistic theory. Meanwhile, one interesting thing in Palmer's definition of collocation in the above quotation is that his definition extends further than many later definitions. In principle, he acknowledges that there is no constraint on the number of constituents of a collocation. This is contrary to the views of Haussmann (1985) and Heid (1994). This will be discussed further later in this chapter under survey of definitions of collocation.

2.3 Theoretical Perspectives on Collocations

This section is devoted to the main theoretical frameworks within which the collocation phenomenon has been addressed in the linguistic literature. This is followed by a survey of definitions of collocations situating them in the theoretical perspectives they represent.

2.3.1 Contextualism

The phenomenon of word collocation has been addressed in the theoretical literature from different perspectives; prominent among them is Contextualism (Firth, 1957; Halliday, 1978). Contextualists reckon that the study of language cannot be done without considering the words' context. Malinowski, one of the key researchers associated with the tradition of 'British Contextualism' argues that "a statement, spoken in real life, is never detached from the situation in which it has been uttered … the utterance has no meaning except in the context of situation" (Malinowski, 1923: 307). What this suggests in the essence is that, meaning of words is defined by their co-occurrence with other words. Right from the early days of collocation research, the concept of word collocation plays a central role in Contextualism. Firth (1957:196) writes about "meaning by collocation" which he defines as "an abstraction at the syntagmatic level […] not directly concerned with conceptual or idea approach to the meaning of the words". Meaning by collocation was first conceived as lexical meaning – one of Firth's five dimensions of meaning (phonetic, lexical,

morphological, syntactic and semantic). As he states, words are "separated in meaning at the collocational level" (1968: 180). Contextualism as one of the theoretical frameworks within which collocations have been described has gone through several stages. It was initially given in terms of habitual co-occurrence of words within a short space of each other in a text (Sinclair, 1991). This 'short space of time in a text' is what Sinclair refers to as collocational span. In a Firthian definition of collocations, the parameter of a recurrent co-occurrence of lexical items translates directly into co-occurrence frequency in a corpus, where the context is usually taken to be a collocational span of 3 to 5 words to either side (Bartsch and Evert, 2014). However, when Sinclair was elaborating further on the framework of Contextualism, he seemed to pay less attention to the distance between collocation items in text. He pointed out that "on some occasions, words appear to be chosen in pairs or groups and these are not necessarily adjacent" (Sinclair, 1991: 115). With this position, the collocating items are not necessarily required to be in the strict proximity of each other.

2.3.2 Text Cohesion

The notion of collocations has also been addressed, though not exhaustively, from the viewpoint of text cohesion. Text cohesion, according to Halliday and Hassan (1976: 4) means "the relations of meanings that exist within text". They distinguish two types of text cohesion namely: grammatical cohesion and lexical cohesion. Collocation is considered an important element of lexical cohesion. From this theoretical standpoint, Halliday and Hassan (1976: 284) see collation as "the association of lexical items that regularly co-occur". Under this theoretical framework, collocation is essentially understood in the same way as in Contextualism. The cohesive effect of collocation is, therefore, derived from words' "tendency to share the same lexical environment" (ibid: 286). According to them, collocations do not only refer to pairs, but also to longer "chains of collocational cohesion" (ibid: 287). They also note "a continuity of lexical meaning" in a collocation through which the cohesion effect is achieved. However, they acknowledge the meaning relations are not easily classifiable in systematic semantic terms. Collocational word similarity is considered a source of text cohesion that is hard to measure and quantify (Kaufmann, 1999). As pointed out earlier, collocations have not been explored exhaustively from this theoretical

perspective. Much of the relevant linguistic literature is heavily reliant on Halliday and Hassan (1976).

2.3.3 Meaning-Text Theory

Collocations also received a formal characterisation within the Meaning-Text Theory (MTT). The Meaning-Text linguistic theory is a theoretical framework for the construction of models of natural language called Meaning-Text Models (Milicevic, 2006). The MTT approach to language was launched in Moscow by Žolkovskij and Mel'čuk in the 1960' and early 1970' (Žolkovskij and Mel'čuk, 1967; Mel'čuk, 1974). The theory places strong emphasis on semantics and considers natural language primarily as a tool for expressing meaning. It is basically interested in linguistic synthesis rather than analysis and has always considered relations rather than classes to be the main organising factor in language. It provides a large and elaborate basis for linguistic description. Within the framework of Meaning-Text Theory is a formal concept called Lexical Function(s). It was first introduced by Žolkovskij and Mel'čuk (1967). Lexical Function is a tool to describe the semantic and syntactic aspects of lexical relations between words in a natural language (Kolesnikova and Gelbukh, 2015).

The tool can be used to describe and systematize two types of lexical phenomena that turn out to be of the same logical nature (Mel'čuk, 1998). The first type of the lexical phenomena involves paradigmatic lexical correlates of a given lexical unit while the second involves syntagmatic lexical correlates of a give lexical unit. This is the one which is particularly relevant to collocation research as it is used to generalize and represent both semantic and syntactic structures of collocations.

2.4 A Survey of Definitions of Collocations

No concept in linguistics seems more variously defined than collocation. Being a borderline phenomenon ranging between lexicon and grammar, it is quite difficult to define and treat systematically. This complexity has given rise to diverse notions of collocation being

propounded by various authors in the last 80 years or so. This disagreement on the notion of collocation is not confined to historical context but also in current research. As Bahns (1993:57) puts it: "collocation is a term used and understood in many different ways". Hence, the term collocation is somewhat often accompanied by confusion, and used in different places to denote different linguistic phenomenona. However, despite the diversity of understandings and points of view, two main perspectives on the notion of collocation can be identified in the literature. These perspectives are 'purely statistically motivated' and 'linguistically motivated' approaches to the definition of collocations (Seretan, 2008). These perspectives are essentially based on five fundamental aspects namely: grammatical boundness, lexical selection, semantic cohesion, language institutionalization, and frequency and recurrence (Pecina, 2010). The 'purely statistically motivated' approaches regard collocations as symmetrical relations and pay no attention to the relative importance of the constituent elements (Seretan, 2008). On the other hand, the syntactic relationship between the constituent elements is a central defining feature of the 'linguistically motivated' approaches to the definition of collocations. The survey of definitions will revolve around these perspectives.

2.4.1 Statistical Approaches

I will start the survey of definitions of collocations with Firth's oft-cited definition of collocation. He observes that:

> "Collocations of a given word are statements of the habitual and customary places of that word" (Firth, 1957: 181).

This Contextualist definition is one of the earlier definitions of collocation. Considering the examples he provided like *night – dark, bright – day, milk – cow* (1957: 196), the understanding he adopted for the notion of collocation seems to be broad. In addition to the syntactic association as in the case of *dark night* and *bright day*, it also covers non-syntagmatic associations which are purely semantically motivated as in the case of *milk – cow*. With the above examples, he claimed that one of the meanings of *night* is its collocability with *dark*, and one of the meanings of *dark* is its collocability with *night*. This suggests that a complete analysis of the meaning of a word would have to include all its collocations. Firth's definition is given exclusively in statistical terms. This statistical view

of collocation is predominant in the work of the so-called Neo-Firthians – Firth's students and disciples – who further developed his theory. They view collocation as the frequent occurrence of one word in the context of another. The context in this case could be the whole sentence or a window of words which Sinclair (1991) refers to as collocational span. The following definitions reflect this view:

> "Collocation is the co-occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening" (Sinclair, 1991:170).

Other definitions which are given exclusively in statistical terms include:

> "The term collocation will be used to refer to sequences of lexical items which habitually co-occur" (Cruse, 1986: 40).
>
> "A collocation is an arbitrary and recurrent word combination" (Benson, 1990).
>
> "Natural languages are full of collocations, recurrent combinations of words that co-occur more often than expected by chance and that correspond to arbitrary word usages" (Smadja, 1993: 143).

In the above definitions, collocation is described in terms of typical co-occurrence or words that show a tendency to occur together. However, they are silent on the syntactic relationship between the constituent elements of collocations. The statistical approaches' view of collocations as symmetrical relations is reflected in Firth's description of collocations in terms of mutual expectation: "the collocation of a word or a 'piece' is not to be regarded as mere juxtaposition; it is an order of mutual expectancy" (Firth, 1968: 181). Cruse also expresses the same view when he concludes that in a collocation "the constituent elements are, to varying degrees, mutually selective" (Cruse, 1986: 40). Sinclair sees collocations in the same light. He describes collocation as "one of the patterns of mutual choices" (Sinclair, 1991: 173).

Halliday (1966), one of the researchers who work within the Neo-Firthian school of thought, defines collocations as "a linear co-occurrence of relationship among lexical items which co-occur". It was Halliday who introduced the term *set* as "the grouping of members with like privilege of occurrence in collocation". For example, words like *hot, bright, shine, light* and *come out* which could collocate with the word *sun* belong to the same lexical set. In a later study, Halliday and Hassan (1967: 287) describe collocation as "a cover term for the cohesion that results from the co-occurrence of lexical items that are in some way or other typically associated with one another, because they tend to occur in similar environment". All the definitions that have been reviewed so far have attempted to capture the essence of

collocations. One thing that is common to all the definitions is that they generally characterised collocations as frequently recurrent co-occurrences of lexical items. The definitions are framed around such notions as frequency, typicality or tendency which are all features usually modelled in statistics. In fact, most of the collocation definitions including the linguistically motivated have elements of statistics in them. The only difference is that the linguistic approaches emphasize the linguistic status of collocations, considering them as well-formed syntactic construction. Consequently, the participating words must be related syntactically (Seretan, 2008).

2.4.2 Linguistic Approaches

The contextualist approaches to the description of collocation seem to ignore the structural relation between items in a collocation. For instance, Sinclair (1991: 170) describes collocation as "lexical co-occurrence, more or less independent of grammatical pattern or positional relationship". On the contrary, the linguistic approaches consider the syntactic relationship between these items as a central defining feature. I will start the survey of definitions of collocations that are based on the linguistic approaches with Cowie – one of the doyens of phraseological approaches to collocation research. He defines collocation as "co-occurrence of two or more lexical items as realizations of structural elements within a given syntactic pattern" (Cowie, 1978:132). This description is consistent with Kjellmer (1987:133) who defines collocation as "a sequence of words that occurs more than once in identical form in a corpus and which is grammatically well structured". What distinguishes these definitions from the statistically motivated ones are the inclusion of "syntactic pattern" and "grammatically well structured" in their description of collocation.

In addition to the above, the linguistic approaches to collocation also address the semantic transparency and opacity of collocations. Laufer and Wildman (2011: 148 – 149) for instance, "regard collocation as habitually occurring lexical combinations that are characterized by restricted co-occurrence of elements and relative transparency of meaning." Restricted co-occurrence distinguishes collocation from free combinations in which the individual words are easily replaceable following rules of grammar. On the other hand, relative semantic transparency of collocation distinguishes them from other word

combinations, particularly, idioms whose meaning is much less transparent than collocations and is very often opaque because it cannot be understood from the words that constitute them.

In order to understand this phenomenon better, consider the collocation: *strong tea,* for instance, which is a restricted co-occurrence. While *strong* can collocate with *tea*, *powerful* which is synonymous to *strong* cannot collocate with *tea*. Looking at the same example from the perspective of 'relative semantic transparency', the collocation *strong tea* is relatively semantically transparent but not fully transparent. The collocate, *strong* has acquired additional meaning. In this context, it means rich in certain ingredients. More examples of relatively semantically transparent collocations include: *heavy drinker*, *strong evidence,* etc.

Another definition which also addresses this aspect of collocation is Chouek (1988 cited in Seretan, 2004:5). He defines collocation as "a sequence of two or more consecutive words that have characteristics of a syntactic and semantic unit whose exact and unambiguous meaning or connotation cannot be derived directly from the meaning or connotation of its components". This means, as Cruse (1986: 40) puts it: "each lexical constituent is also a semantic constituent". Each lexeme makes an independent contribution to the meaning of the whole collocation. This independent meaning of constituents marks off non-idiomatic combinations from idiomatic expressions and this differentiates collocations, in the narrow sense of it, from other lexical, non-idiomatic combinations (Trantescu, 2015).

In continuation of the survey of the definition of collocation, I will consider a few more definitions which are based on the linguistic approaches to collocation. One of such definitions is given by Bartsch (2004). She defines collocation as "lexically and/or pragmatically constrained recurrent co-occurence of at least two lexical items which are in a direct syntactic relation with each other" (ibid: 76). This definition regards collocation as a syntactically-bound word association. This syntactic well-formedness criterion implies that the collocational span is the phrase, clause or the sentence containing these words. All the definitions that have been considered so far – both the statistical and the linguistic approaches – have one thing in common which is the recurrence of the phenomenon. This recurrence is maintained as a defining feature, and this is expressed by such attributes as "conventional",

"recurrent", and "characteristic". Furthermore, collocations have been viewed as a directed relation in which the role played by the constituting elements is uneven (Halliday, 1966) which means collocations have hierarchical structuring. The node, also known as the collocational base, and the collocate are in a directed relationship. What is meant by directed relationship is that the collocational base (node) collocates with the collocate and not vice versa. In this directed relationship, the collocate further specifies the meaning of the collocational base.

2.5 The Core Defining Criteria of Collocations

A review of the literature has revealed a multitude of collocation definitions which are quite divergent. This divergence of definitions may lead to confusion despite the fact that a clear distinction can be drawn based on the underpinning approach (linguistic or statistical approach). This section is aimed at identifying the core defining features of collocations. These defining features are the ones that are more recurrently mentioned, and which appear to be accepted by most collocation researchers. These features are a kind of point of convergence for most of the authors who have tried to define the collocation concept.

A review of the key criteria commonly deployed in defining collocation in the research literature is necessary. This is to provide a clearer picture of which of these criteria should be employed in the identification and characterisation of collocations in this corpus study and why. One criterion that features prominently in most definitions of collocation, particularly, the more statistically inclined definitions is frequency of co-occurrence (Benson et al, 1986; Kjellmer, 1987; Smadja, 1993). Computer-aided corpus studies have revealed much more reliably than native speaker intuition that many words in the English language have tendency to recur in combination with a very limited number of other lexical items. The frequency of co-occurrence of particular word combinations within the same immediate context is an empirically verifiable feature of collocation (Bartsch, 2004). So, the following are the core defining criteria of collocations:

2.5.1 Collocations are Prefabricated Phrases

They are available to speakers as ready-made or prefabricated units. They contribute to fluency and naturalness of speakers' utterance (Pawley & Syder, 1983). We acquire collocations as we acquire other aspects of language through encountering texts in the course of our lives (Hoey, 2000). In Sinclair's words, the language is governed by two opposing principles namely: the open principle and idiom principle. The open principle refers to the regular choice in language production while the idiom principle refers to the use of prefabricated units which are already available. Collocations belong to the idiom principle. Sinclair (1991: 110) refers to collocations as "semi-prefabricated phrase that constitute single choice even though they might appear to be analysable into segments". The idea of collocations as prefabricated unit has earlier been expressed by Palmer (1938) and Hausmann (1985). Palmer refers to collocations as "words that may be best learnt as if it were a single word". The reference to collocation as 'a single word' suggests that collocation is prefabricated and could be acquired and used as one chunk. In the same vein, Hausmann (1985: 124) calls them "semi-finished products" of language.

2.5.2 Collocations are Arbitrary

Several definitions of collocation in the literature refer to the arbitrariness of collocations. They are not regarded as regular productions of language, but rather "arbitrary word usages" (Smadja, 1993), "arbitrary […] word combinations" (Benson 1990), or as Hausmann (1985) puts it, "a typical, specific and characteristic combination of two words". Other major definitions that take note of this feature include: Fontenelle (1992) and van der Wouden (1997). Fontenelle refers to collocations as "idiosyncratic syntagmatic combination of lexical item" (Fontenelle, 1992: 222) while van der Wouden (1997) refers to them as "idiosyncratic restriction on the combinability of lexical items".

The fact that collocations are prefabricated units in the lexicon of a language suggests that they are to be acquired and used as such. This will, therefore, prevent the reconstruction of collocations by means of grammatical process. The arbitrary nature of collocation means it is difficult to explain the reason for a particular choice of words in a collocation simply based

on the rule of grammar and syntax. On the contrary, it seems once this choice was made and conventionalized or institutionalized, using Sig's (Seg et al, 2002) term, other paraphrases (of such combinations) are blocked as specified by Sinclair's idiom principle. The arbitrariness of collocation is not limited to the choice of a particular word in conjunction with another in order to express a given meaning as Kahane and Polguere pointed out (Kahane and Polguere, 2001 cited in Seretan, 2004). But it is also arbitrary in terms of its syntactic and semantic properties. According to Evert (2004: 17), "collocation is a word combination whose semantic and/or syntactic properties cannot be fully produced from those of its components, and which therefore has to be listed in a lexicon".

2.5.3 Collocations are Unpredictable

One of the reasons why collocation is notoriously difficult to acquire and produce by second language learners is that, "the affinity of a word for a particular collocate which is strongly preferred over other words from the same synonymy set is unpredictable" (Seretan, 2004: 16). This unpredictability is another main feature that is often cited in collocation definitions. Evert (2004: 17) states that the "syntactic properties (of collocations) cannot be fully predicted from those of its components". This is so because the 'institutionalization' of a collocation as a prefabricated unit does not seem to depend on clear linguistic reasons. It is not possible to predict the morpho-syntactic properties of a collocation on the basis of the properties of the participating words (Seretan, 2004). According to Cruse (1986), the affinities between the constituents of a collocation cannot be predicted on the basis of semantic or syntactic rules, but rather can only be observed with some regularity in text. As a result of this arbitrariness, collocation is not reproducible by simply applying the grammatical prescription of a language.

2.5.4 Collocations are Recurrent

This is the feature of collocation that is mostly remarked in the various definitions in the literature. Collocations are "habitual and customary" (Firth, 1957:181), they are "actual words in habitual company" (Firth, 1968: 182). In the words of Benson (1990), they are "combinations of words that co-occur more often than expected by chance". Collocation is

undoubtedly recurrent in language. It is their frequent usage that determines their 'institutionalisation'. It is the same frequency of usage that makes them "psychologically salient" (Benson et al, 1986b: 252). If not for their frequency, we would probably not have recognized them.

2.5.5 Collocations are made up of two or more words

Although collocation research in the literature is almost exclusively concerned with collocations made up of two lexemes, theoretically, there is no length limitation for collocations**.** This is further stressed by Sinclair (1991: 170) who points out that "in most of the examples, collocation patterns are restricted to pairs of words, but there is no theoretical restriction to the number of words involved". In actual fact, a vast majority of the definitions specify that collocation is "the co-occurrence of two or more words within a short space of other" (Sinclair, 1991: 170), "sequence of two or more consecutive words" (Choueka, 1988 cited in Seretan, 2004: 16); "co-occurrence of two or more lexical items" (Cowie, 1978). Examples of collocations that have more than two lexemes are: *abolish the death penalty, major turning point* and *conduct a comprehensive study* (Seretan, 2004).

2.6 Classification of Collocations

Collocations are considered a type of word combination in certain grammatical pattern which means the term 'collocation' will be used both to refer to an abstract unit of language and its instantiations in texts. Three major types of classifications of collocations can be identified in the literature. One type, which is the most comprehensive of them, is based on the syntactic characteristics of the collocation. Another one is based on the semantic characteristics while the third is based on the commutability of its element. Commutability means the substitutionability of the constituents of a collocation with their synonyms.

Haussmann (1989) classifies restricted collocations based on the syntactic characteristics of the constituents. He classifies them according to the word classes their constituents belong. He divides collocations into six types namely: adjective + noun, noun + verb, noun + noun, adverb + adjective, verb + adverb, and verb + noun. Aisenstadt (1981) has earlier proposed a similar classification; however, she divides the verb + noun group further into verb + noun and verb + prep + noun. Benson et al (1986) also make the same classification as Haussmann but added the combination noun + prep, prep + noun and adjective + prep. This is probably because of the broader nature of their definition of collocation. They went further to make more basic distinction on the ground of the word classes to which the constituents of the collocation belong. They call collocations in which two lexical items occur as "lexical collocations" while collocations in which lexical and more grammatical elements co-occur are called "grammatical collocations". Most studies in the literature use the BBI classification of collocations. The BBI classification divides lexical collocations into seven groups (which is similar to Haussmann's apart from the verb + noun). Grammatical collocation on the other hand, is divided into eight groups namely: G1 – G8, with G8 further divided into nineteen sub-groups. G1 is noun + prep, G2 is noun + to – inf, G3 noun + that – clause, G4 is prep + noun, G5 is adj + prep, G6 is pred adj + to – inf, G7 is adj + that – clause, and G8 is verb + various grammatical pattern/combinations.

The second type of classification of collocation is based not on the syntactic characteristics of the combination, but purely on the semantic characteristics of what Haussmann (1989, cited in Nesselhauf, 2004: 22) calls the 'collocator'. Cowie (1992) also attempts to classify collocation this way though limited to the verb + noun collocations. He distinguishes between verbs with "figurative, delexical and technical or semi-technical" meaning (Cowie, 1992: 5). Example of a collocation with a delexical verb is 'make proposal'; the one with a figurative verb are 'dismiss the suggestion' 'abandon a principle' and the one with technical or semi-technical verb are 'enact measures' 'draft the legislation'. Cowie's classification is not as detailed as Mel'čuk's who also classified collocation on the basis of lexical function. Lexical functions describe the combinatorial properties of lexical units. As a concept, it was introduced within the framework of the Meaning-Text Theory (Mel'čuk, 1974, 1996) in order to explain the lexical restrictions and preferences of words in choosing their 'companions' when expressing certain meaning in text (Gelbukh and Kolesnikova, 2013). A lexical function is a meaning that may be expressed by a variety of different lexemes. What this

means is that in a given collocation, the lexeme(s) which expresses this meaning is chosen by the keyword. This keyword is referred to as the 'base' (Haussmann, 1984) or the 'node' in Halliday's (1966) term. The base is semantically autonomous and the collocate needs the base in order to get its full meaning. Having reviewed the existing literature on the general phenomenon of collocation, tracing its establishment and development in linguistic theory as well as the main theoretical frameworks within which the concept of collocation has been addressed so far in the linguistic literature, I will now focus on studies on L2 collocations.

2.7 L2 Collocational Research: state of the art

It has been over three decades since Pawley and Syder (1983: 191) discussed their "two puzzles for linguistic theory: nativelike selection and nativelike fluency." Their study focused on two issues. The first was on "the ability of the native speaker routinely to convey his meaning by an expression that is not only grammatical but also nativelike … natural and idiomatic from among the range of grammatically correct paraphrases, many of which are non-nativelike or highly marked usages." (ibid). The second is "the native speaker's ability to produce fluent stretches of spontaneous connected discourse … [the puzzle of the] capacities for encoding novel speech in advance" (ibid). They were particularly interested in "the features that make certain forms of expression 99 per cent more likely to occur in a given everyday context than their paraphrases, which are equally grammatical" (ibid: 199). Over the years, we have come to understand that non-native speakers, even advanced speakers, have limited ability to produce expressions that are nativelike, natural and idiomatic from a range of grammatically correct paraphrases – expression which are 99% more likely to occur in a given everyday context than their paraphrases, which are equally grammatical. The expressions being referred to here are obviously formulaic expressions like collocations. Three decades on, as Wray (2012: 23) rightly notes, "something about formulaicity as a property of language has captured researchers' imagination, and there seems to have been an explosion of activity" in the last two decade or so. There has, particularly, been an increasing interest in L2 collocations research. The focus of this section is two-fold: (1) to review the existing literature on L2 learners' collocational competence and development from around

the world. This review will be limited to studies that are most relevant to my study. (2) to review the few studies on collocations in Nigeria.

The often-cited earlier research on collocations by Biskup (1992), Bahns and Eldaw (1993), Bahns (1993) Lewis (1993), and Farghal and Obiedat (1995) seem to have drawn language teachers and researchers' attention to the frequency and importance of formulaic sequence in both language learning and language use as well as the difficulties learners have producing them. Of all the formulaic sequences, collocations have received much attention. A survey of the existing literature reveals, among other things, that various studies have investigated L2 learners' collocational competence and development, cross-linguistic influence (L2 negative transfer) on the production of collocations (Leśniewska & Witalisz, 2007; Shehata, 208; Yamashita & Jiang, 2010; Phoocharoensil, 2012), and L2 collocations receptive knowledge (Nizonkiza, 2015; Begagić, 2015). There has also been a focus on the effect of congruency, frequency of input, and immersion on collocational knowledge. Other studies have identified, classified and analysed collocational errors, investigated collocational processing and explored the relationship between proficiency and collocational knowledge. A substantial number of studies have also investigated the teaching of collocations looking at the impact of various teaching and learning approaches on L2 collocations. Some of the issues raised in these studies will be addressed in this section.

Based on Kroll and Stewart's (1994) Revised Hierarchical Model and Jiang's (2000) model of L2 mental lexicon, Yamashita and Jiang (2010) investigate the influence of L1 on the acquisition of L2 collocations. They compare the performance on a phrase-acceptability judgment task among L1 English speakers, 24 Japanese English as a second language (ESL) users, and 23 Japanese English as a foreign language (EFL) learners. The ESL group were Japanese students, researchers or instructors residing in the US as at the time of the study while the EFL were Japanese residing in Japan who had never lived in English-speaking country. It is not clear though how long they have resided in the US. Their findings indicate that the EFL group "made more errors with and reacted more slowly to incongruent collocations than congruent collocations" while the ESL group generally performed better making fewer errors and responded faster although they too made more errors on incongruent collocations than on congruent collocations. However, L1 influence was not apparent on the

ESL groups' reaction time. They conclude "both L1 congruency and L2 exposure affect the acquisition of L2 collocations with the availability of both maximizing this acquisition" and that the acquisition of incongruent collocations is difficult even with a considerable amount of exposure to L2. Their conclusion is consistent with Groom (2009) who also concludes that the acquisition of collocation is difficult regardless of the amount of exposure.

Meanwhile, Kroll and Stewart's (1994) Revised Hierarchical Model (RHM) of bilingual language processing is theoretically central to this study and will be used to explain some of the findings of this study in the discussion chapter. The model essentially merged the word association and the concept mediation models (Potter et al, 1984) into one single developmental model. The word association model proposes that a direct association is established between words in the two languages and that this association is used to understand and produce words in the L2 by retrieving a word in the L1 in the course of second language acquisition. The concept mediation model, on the other hand, proposes that "the only connection between the two languages is via an underlying, amodal conceptual system" (ibid: 23). The RHM makes a hierarchical distinction between two types of word representations – lexical representations containing information about word forms and the conceptual representations corresponding to the word meanings. Two lexicons are distinguished at the lexical level – one for words of the L1 and one for the words of the known L2. And there are excitatory connections between translation equivalents at the lexical level. These connections are assumed to be much stronger from L2 to L1 particularly at the early stages of language proficiency because many L2 words are learned by associating them with their L1 translation equivalents. Besides, the L1 lexicon is larger than the L2 lexicon. This suggests L2 learners may easily produce congruent words (words that could be associated with L1 translation equivalents) but have difficulty producing incongruent words (words that could not be associated with L1 translation equivalents).

The two lexicons – the L1 and L2 lexicons – are connected to a shared conceptual system that contains the meaning of the two words. Both the lexical and conceptual links are active in the bilingual memory according to the RHM. However, the strengths of the links differ depending on fluency in L2 and comparative dominance of L1 to L2. At the conceptual level, the model assumes a direct connection from the L2 word to its conceptual

representation. Kroll and Stewart (1994) point out that the links between the L1 words and the conceptual system are stronger than those between the L2 words and the conceptual system. For someone who learns L2 beyond a stage of very early childhood, there would have been a strong link between their L1 lexicon and their conceptual memory. At the initial stages of L2 learning, the L2 words are linked to this system by lexical links with their L1. However, as they become more proficient in the L2, direct conceptual links are also required but the lexical links do not disappear when the conceptual links are established.

Returning to the review of the studies on L2 collocations, in a study similar to Yamashita and Jiang (2010), Shehata (2008) studies two groups of Arabic-speaking learners of English – one group consists of 65 university students in the US which she categorises as ESL and the other consists of 62 undergraduate English major students in Egypt which she categorises as EFL. She uses a combination of questionnaire, gap-filling tests, appropriateness judgment test, and vocabulary recognition test to explore the impact of learning environment and exposure to the target language on the acquisition of collocations. Her findings show the ESL group performs better than the EFL group which suggest learning environment strongly influence the acquisition of collocation. She interprets this as a positive correlation between collocational knowledge and exposure to target language. She also finds evidence of L1's influence on collocational knowledge with the learners having difficulty with incongruent collocations. She concludes that the learners' productive knowledge of collocations lags behind their receptive knowledge. Her findings are consisting with Yamashita and Jiang (2010) above.

Various studies have shown L1 transfer as being common in L2 collocations acquisition but most of them did not indicate at what level the L1 transfer occurs. Song and Wolter (2017) study this phenomenon a step further by investigating whether L1 transfer occurs in L2 verb noun collocational production at the semantic preference and semantic prosody levels. They conduct cross-linguistic comparisons to explore the different semantic preference and features between ten high frequency English verbs and their Chinese equivalents and to determine whether the cross-linguistic semantic differences have effect on L2 learners' collocational output. They use data from three corpora: the Corpus of Contemporary American English (COCA), the Beijing Language and Culture University Chinese Corpus (BCC), and the English Compositions of Chinese Learners Corpus (TECCL). Using data

from COCA and BCC to establish degree of overlap for semantic preference between translation equivalents of verbs in English and Chinese, they conduct cross-linguistic comparisons to explore the different semantic preference and features between ten frequently used English verbs and their Chinese equivalents to determine whether the cross-linguistic semantic differences have effect on the L2 learners' collocational output. Their findings suggest that the tendency of L2 learners producing native-like collocations is strongest where semantic preference overlap between the English verbs and their Chinese equivalents which means L1 transfer occurs, as they put it, "not only at the semantic and syntactic level, but also at the collocational level … semantic preference features of a verb stored in one's L1 mental lexicon were also activated and in effect in L2 learners' VN collocational output process" (ibid: 1). They, however, find semantic prosody values to be less reliable in predicting native-like collocations.

The relationship between proficiency and collocational knowledge has attracted much interest in the last decade or so. All the studies in this area seem to suggest collocational knowledge increases in consonance with proficiency increase (Hsu & Chiu, 2008; Nizonkiza, 2012, 2015). Nizonkiza (2012) investigates the relationship between productive knowledge of collocations and academic literacy among first year students at North-West University, South Africa. Using items selected from Nation's (2006) word frequency bands the Academic Word List (Coxhead, 2000), he administers a collocations test on the participants. His findings indicate that collocational knowledge correlates with academic literacy. Some years later, he investigates receptive collocational competence across proficiency levels (Nizonkiza, 2015). His findings indicate that receptive collocational knowledge develops alongside proficiency. This lends empirical support to Hsu and Chiu's (2008) study of the relationship between the production of collocations and speaking proficiency in Taiwan. Their findings suggest that the learners' knowledge of lexical collocations correlates with their speaking proficiency. All these findings are consistent with the findings of other studies on the link between collocational competence and linguistic proficiency (Laufer & Waldman, 2011).

The literature is awash with studies that analyse collocational errors. The nature of collocational errors that learners make seems to be the most studied aspect of L2 collocations. Most of these studies have focused on identifying, classifying and analysing the errors

(Farghal & Obiedat, 1995; Shih, 2000; Boonyasaquan, 2009; Phoocharoensil, 2011). The error analyses in the literature are predominantly focused on the influence of L1 which seems to be the greatest source of collocational errors (Chen, 2004; Nesselhauf, 2005; Hama, 2010; Laufer & Waldman, 2011). The drive to investigate collocational errors seems to have led to a dearth of studies on how much collocations learners know. While a better understanding of the nature and causes of collocational errors is important, particularly for language pedagogy, the overwhelming focus on learners' collocational deficiency seems to have created an impression that L2 learners do not have much collocational knowledge. Very few studies have pointed out that learners could produce substantial numbers of well-formed collocations (Fernández & Schmitt, 2015). Most of the corpus-based collocational error analyses have often used their reference corpus as a baseline for determining unacceptable collocational combinations without considering how many well-formed collocations the learners produce in comparison to the native speakers. One of the few exceptions is Laufer and Waldman (2011) who compared collocations in Israeli Learner Corpus of English with LOCNESS – a native speaker corpus. Besides, most of the error analyses did not go beyond merely pointing out the errors without an in-depth linguistic analysis of the errors in developmental terms in a way that could account for L2 mental lexicon.

Finally, in this section, there is a gap in the literature in terms of studies of collocations from the Nativized Englishes. There is also a dearth of studies that explore the structural and semantic properties of collocations produced by L2 learners. To the best of my knowledge, there is no study in the literature that analyses the structural and semantic properties of the constituents of the well-formed collocations produced by learners in comparison to native speakers. There seems to be a neglect of the influence of the semantic properties of collocations on the collocational knowledge of learners. Some aspects of this gap in the literature are addressed in this thesis. Meanwhile, let us now focus this review on studies on collocations in Nigeria.

2.7.1 L2 Collocational Research in Nigeria

While collocations have received much attention in the last 20 years or so, it has not been the focus of many studies in Nigeria. This is despite their frequency in language and the importance of the mastery of collocations as being central to communicative competence (Barfield & Gyllstad, 2009b; Schmitt, 2004). Until recently, collocations did not feature in most of the English language teaching textbooks in Nigeria. Even now, collocations have only received marginal attention in the textbooks. This probably explains the dearth of collocational research in Nigeria. By implication, this may be indicative of lack of awareness of both the problem collocations pose to L2 learners and the importance of collocational competence as facilitator of fluency. Besides lack of awareness, the apparent dearth of interest in collocational research may also be due to limited access to research instruments such as corpora and corpus analysis tools.

In one of the earliest publications on collocations, Taiwo (2004) writing on the importance of collocations in English as second language acquisition, stresses that the neglect of collocations in Nigerian English curriculum should be a concern for teachers. He notes that much of the language research efforts in Nigeria are being concentrated on the grammatical, phonological and orthographical levels at the expense of the lexical levels. Writing as a member of the English language teaching community in Nigeria, he observes that where the lexical aspect is taught at all, teachers prioritise the paradigmatic sense relations of lexical items at the expense of collocations. His observation reflects the neglect of collocations in the textbooks. Some years earlier, Taiwo (2001) analyses 200 letters written by 15 – 20 years old Yoruba-speaking, final year students from ten randomly selected secondary schools. He identifies a total of 85 lexico-semantic relation errors out of which 48 representing 56.6% are collocational errors. He finds out that the learners fail to observe the rule of restrictions on the co-occurrence of lexical items resulting in collocational errors such as substitution of collocates with their synonyms, clang association among other things. While he classifies the collocational errors, he however, neither attempts to provide linguistic reasons for these errors nor discusses the proficiency levels of the participants.

Okoro (2013) explores collocational usage in Nigerian English to discover their structural composition and pattern of errors. He got his data from various sources including "spoken

usages overheard among Nigerians of all persuasions in all sorts of formal and informal contexts ..., secondary sources documented in the literature on Nigerian English" (ibid: 97); texts from unidentified students' essays, print media and textbooks; and his own retrospection as a speaker of Nigerian English. His structural analysis of the collocations and the patterns of the collocational errors in the texts reveal omission of collocational elements, the inclusion of redundant collocational elements, and the substitution of the lexical element in some collocations. He also discovers the restructuring of collocations which results in infelicitous combinations and the alteration of the grammatical property of collocational items.

Okoro's findings seem to suggest collocational deficiency is pervasive in Nigeria. However, the data he used, his concept of collocations and some of his claims seem problematic. Some of his data are texts from unidentified students' essays, print media and textbooks. The data is not clearly defined in terms of the English proficiency. Collocational competence has been found to be much related to general language proficiency (Hosseini & Akbarian, 2007; Namvar, 2012; Ebrahimi-Bazzaz, et al, 2014). Not defining the proficiency level the texts represent makes the findings of little value. Besides, his concept of collocation seems not properly delineated as some of the examples of collocations he provided are completely different from the various examples in the existing literature on collocations. The examples of the collocations he provides such as: *'for one good year', 'sitting behind the steering wheel', 'Sauce for the goose is sauce for the gander'* do not fit in to any of the definitions of collocations in the existing literature. The third example (Sauce for the goose is sauce for the gander) is an idiom rather than a collocation. Furthermore, he claims that one "unique feature of collocational usage in Nigerian English ... is marked absence of many of the collocations which are common in native-English usage" (Okoro, 2013: 109). This claim is unsubstantiated as he did not provide any frequency data or compares his results with any Native English corpus. He also identifies some collocations which he describes as being peculiar to Nigeria but then regard them as "obviously sub-standard" (ibid: 111). It is not clear why he regards certain collocations which may be variety marker of Nigerian English as sub-standard. Perhaps, his notion of acceptable collocation is based on the norms and standards of any of the prestigious varieties of English. I will explore the issue of norms further in the discussion chapter.

In another study, Shittu (2015) investigates collocational errors in the essays written by students of a Federal College of Education Norther Nigeria. She regards the learners as

advanced speakers of English though most people in Nigeria will not regard students of College of Education as advanced speakers of English. According to her, all the participants are multilingual and of similar language proficiency. It is not clear how she determines their language proficiency. But there is no evidence she carefully defines the language proficiency of the participants. This seems to be a common problem with the few studies on collocation in Nigeria. By crudely labelling the population of their studies as 'advanced' or "Nigerians of all persuasions in all sorts of formal and informal contexts" (Okoro, 2013: 97) means little in developmental terms as Hulstijn et al. (2010) point out. Notwithstanding this apparent shortcoming, Shittu's study reveals the participants had difficulty producing collocations. Most of the errors she identifies were mainly L1-induced and overgeneralisation. Her conclusion was that "students' collocation errors are attributable to poor teaching and learning which resulted in wrong generalization of rules" (Shittu, 2015: 3176). She did not present any evidence to substantiate this claim.

Israel (2014) investigates lexico-syntactic errors in teaching materials (textbooks) written by bilingual Nigerian authors who had their education in Nigeria. The errors he identifies include the alteration of grammatical properties in collocational items and substitution of lexical elements within collocational structures. He made a stark conclusion that "students are merely the conveyor belt of errors contained in the teaching material[s]" (ibid: 75). He essentially blames teaching materials and by extension their writers for Nigerian students' collocational deficiency.

Friday-Òtún and Ọmọ́léwu (2016) who are teachers and speakers of Yorùbá conducted a rare research on collocations in Yorùbá language. All the collocational studies in Nigeria have focused on the English language, but their study attempted to describe the structures and types of collocations in the Yorùbá language usage. They extracted collocations from 19 randomly selected examination scripts on two Yorùbá language modules written by L1 Yorùbá University students. The two modules are *Mofoloji Yorùbá* (Yorùbá Morphology); and *Awon Ariyanjiyan tó N Lo ní Abala Síntásì* (Issues in Syntax). They identified three classifications of word combinations which are related to collocations: free combinations, restricted co-occurrences and fixed collocates. This is similar to the English language phraseological units (see Aisenstadt, 1979; Cowie, 1981; Howarth, 1996, 1998). The main source of the collocational errors identified in their study is negative transfer from L2

English. This is seemingly in direct opposite of the findings from various L2 collocational research where the main source of errors is L1-based. This result highlights the influence of the English language on Nigerian indigenous languages. It means the production of L1 collocations could be problematic in certain contexts.

One trend can be identified in the collocational studies in Nigeria. They all focus on identification and classification of errors. They have all manually extracted collocations they regarded as errors from relatively some texts. The scope and depth of these studies are quite narrow leaving much unknown about the collocational competence and development of Nigerian learners of English. Besides, as the literature clearly reveals, the various studies on collocations in Nigeria did not clearly define the English language proficiency of their subjects making their findings to mean little if anything in developmental terms. This is the wide gap in the literature which my study attempts to fill. This study is the first computer corpus-based study of collocations in Nigeria.

Meanwhile, it is important at this stage to provide my definition of collocation. Having reviewed the existing literature on the phenomenon of collocation and a survey of the definitions of collocation as well as reviewing the literature on L2 collocation research, I will adopt a hybrid approach, as I have stated earlier – a midway between the phraseological approach and frequency-based approach of defining collocations. Collocation in this study, is therefore, defined as words that habitually appear together within a given word span, relatively fixed, and thereby convey meaning by association with varying degree of transparency in meaning e.g. *crystal clear, excruciating pain, commit suicide, strong tea, proffer solution* (*Proffer solution* is a Nigerian English Collocation). I will now review the literature on Learner Corpus Research.

2.8 Learner Corpus Research: state of the art

Learner Corpus Research, as a field of scientific enquiry, has grown rapidly within its relatively short existence. Since its emergence in the late 1980s, LCR has been the focus of much active international work (Granger, 2004). Leech (1992: 106) sees its potentials right early when he describes it as "a new research enterprise, a new way of thinking about learner language, which is challenging some of our most-deeply rooted ideas about learner language." It has offered us a new tool for better analysis and understanding of learner language. This brief review will focus on corpus data collection, corpus annotation, learner corpus typology and a survey of learner corpora.

Learner corpus can be categorised as natural or authentic language use data gathered to describe learner language (Granger, 1998, 2004). Learner corpus is very import because it provides a deviation from the standard or native variety of a particular language (Pravec, 2002). Through the study of authentic natural learner language data, we can focus on theoretical and pedagogical issues as well as focus on L2 learners' needs. Because corpus data are stored electronically which means we can quickly and with relative ease, collect large amount of texts, the sizes of learner corpora are becoming bigger – now in the millions. This also means having access to a large amount of learner language in a way that was not possible until the advent of computer corpora. Does this necessarily mean big is better? Bigger might be better, although it depends on the nature of the research. If a corpus is too small, it might not be representative of the target group and this may raise questions on validity of any findings based on the analysis of such corpus. MacWhinney (2000: 3) notes that "conducting an analysis on a small and unrepresentative sample may lead to incorrect conclusions." This concern was further highlighted by Gass and Selinker (2001: 31) when they pointed out that it was "difficult to know with any degree of certainty whether the results obtained are applicable only to the one or two learners studied, or whether they are indeed characteristic of a wide range of subjects." A bigger corpus would be representative enough to be able to generalise results. We do not know for certain how big a corpus needs to be for general or specific purposes. But for the corpus data to be representative sample of the target group, it will have to be fairly big. While it seems the bigger the corpus the better, Kennedy (2014: 68) cautions that "rather than focusing so strongly on the quantity of data in a corpus, compilers and analysts need always to bear in mind that the quality of the data they work

with is at least as important." As Granger (2004: 125) rightly points out, large corpus "is a major asset in terms of representativeness of the data and generalizability of the results", but the size should not be prioritised over the quality of the corpus texts.

To ensure that a corpus contains all the relevant design parameters in terms of the size and the quality of the corpus texts, Biber (1993: 256) suggests that a "theoretical research should always precede the initial design and general compilation of texts." Such research is important because learner language can be influenced by a wide range of factors. These factors include linguistic, psycholinguistic and situational factors (Granger, 2004). Failure to control these factors has potential to limit the validity of any findings on such learner language. Learner corpora are compiled according to strict design criteria (Tono, 2003; Glaznieks et al, 2014) with some of these criteria being the same as for native corpora (Atkins, Clear and Ostler, 1992). What this means is that some randomly collected heterogeneous learner texts would not qualify as learner corpus.

Learner corpus has some functionalities such as count, sort, compare and annotate which lend themselves so well to automation and these functionalities make learner corpus attractive to second language acquisition and foreign/second language teaching research. The count functionality allows for comparison of the frequency of linguistic items in learner corpus texts as well as making it possible for research to get precise figure using the word count option of corpus analysis tools (Granger, 2002). Using the Concordance (sorting), L2 researchers can have a view of the lexico-grammatical pattering of the words produced by the learners. This is one of the reasons why corpus-based method is popular in L2 collocations research. It is also possible to compare learner text with native speaker text as well as two or more L2 texts. Annotation, which Garside et al (1997:2) define as "the practice of adding interpretative, linguistic information to an electronic corpus of spoken and/or written data" can provide researchers additional layers of information which may help with the analysis of the corpus data. Corpus annotation could be necessary in order to test a particular theory (Anthony, 2013).

Granger (2004: 128) points out that "any type of annotation is potentially useful (discourse annotation, semantic annotation, refined syntactic annotation, etc.)" particularly error annotation for interlanguage studies. There, however, exists in the literature argument against corpus annotation. Sinclair (2004b: 191cited in Anthony, 2013) argues that:

> "interspersing of tags in a language texts is a perilous activity, because the text thereby loses its integrity, and no matter how careful one is the original text cannot be retrieved...In corpus-driven linguistics you do not use pre-tagged text, but you process the raw text directly and then the patterns of this uncontaminated text are able to be observed."

It seems Sinclair's objection to corpus annotation is only relevant to researchers who adopt corpus-driven approach. But over the years, there have been various development of corpus analysis tools. As Anthony (2013) points out, there are now corpus analysis tools that are able to show or hide the annotations if the researchers want to analyse raw data. It should be noted though that most of, if not all, the existing corpus annotation programmes are designed on the basis of native speaker corpora and as such may not perform accurately when confronted with learner corpora (Granger, 2004). They have been found to be highly sensitive to morpho-syntactic and orthographic errors (Van Rooy and Schäfer 2003). This means they may not be suitable for automatic tagging of least proficient learner texts which may contain many learner errors. Finally, on this, careful annotation of corpora (including manual verification to correct where tagging tools have made mistakes due to the influence of learner errors), are indeed useful depending on the aims of the corpus analysis and the approach adopted. I will now focus on corpus typology.

Learner corpora in the literature have been classified along the line of longitudinal versus cross-sectional, spoken versus written and commercial versus academic. Longitudinal learner corpora contain texts collected from the same learners over a period of time while cross-sectional corpora contain texts collected from different categories of learners at a single point in time (Granger, 2004). Researchers interested in interlingual development have either used longitudinal corpora or what Granger (ibid: 131) calls "quasi-longitudinal corpora." These corpora contain text collected from learners at the different proficiency levels at a single point in time. Overwhelming majority of the corpora in the learner corpus research literature is cross-sectional and this is followed by quasi-longitudinal corpora. There are still relatively few longitudinal corpora in the literature in comparison to the other types of corpora. In the recent years, there has been an increase in the numbers of longitudinal corpora (Roy, Frank & Roy, 2009; Kumar et al, 2015).

Learner corpus research is dominated by written corpora. This is obviously because the time and effort involved collecting and transcribing spoken corpus data is prohibitive. A vast majority of the learners represented in the learner corpus research are learners of English as a Foreign language (EFL) as opposed to English as a Second Language (ESL) and almost all the learner corpora are in Europe and Asia (Pravec, 2002; Granger, 2004). The terms EFL and ESL are problematic because of the different meanings various researchers have ascribed to them, and sometimes they are used interchangeably. Some of the early usage of these terms used English as a Foreign Language to "mean English taught as a school subject or on an adult level solely for the purpose of giving the student a foreign language competence" while English as a Second Language is used to refer "to a situation where English becomes a language of instruction in the schools, as in the Philippines, or a lingua franca between speakers of widely diverse languages, as in India" (Marckwardt, 1963:25).

However, Granger (2002) situates non-native varieties of English within three categories namely: English as an Official Language (EOL), English as a Second Language (ESL) and English as a Foreign Language (EFL). EOL is "cover term for indigenized or nativized varieties of English, such as Nigerian English or Indian English" (ibid: 5) which seems to be the variety Marckwardt (1963) referred to as ESL. In Gass and Selinker's 2001: 5) view, ESL takes place in a context "with considerable access to speakers of the language being learned, whereas learning in a foreign language environment does not." But I use ESL to refer to a context where there is a nativized variety English and where English is used in everyday life in addition to the local languages as in the formal British colonies like Nigeria, Ghana, India, etc. I use EFL on the other hand to refer to a context where English is not a commonly used language like in China, Brazil, Russia, etc. The conclusion that the existing learner corpora predominately represent EFL is based on the above definition of non-native varieties of English. What is striking in learner corpora research is the dearth of studies on the nativized varieties of English. Computer learner corpus would be a versatile tool for linguistic comparative analysis of interlanguage of the various varieties of World Englishes. Corpus-based studies of these new Englishes will lead to a better description of the various varieties. But there is a pronounced gap in the literature in this area.

A survey of the existing learner corpora shows there is only one learner corpus of English from Nigeria – a country with the largest population of speakers of nativized English after

India. This is not limited to Nigeria; the whole of the new Englishes in Africa is largely unexplored from the perspective of computer learner corpora research. While there could be a few learner corpora used for small scale studies by individual researchers in Africa, there are no open access learner corpora in Africa, at least to the best of my knowledge. There are however ongoing corpus compilation projects such as the Spoken Xhosa English (de Klerk, 2002; 2006) and the Corpus of South African English at the Rhodes University (both in South Africa). But the descriptions of these corpora suggest they cannot be categorised as learner corpora. A survey of the existing learner corpora in the literature shows the extent to which Africa is lagging behind in computer learner corpora research.

A team led by Sylviane Granger at the University of Louvain, Belgium has been at the forefront of learner corpora research. They have developed two of the largest existing learner corpora - The International Corpus of Learner English (ICLE) and Louvain International Database of Spoken Interlanguage Database (LINDSEI). ICLE is a collaborative project with various partner Universities. The corpus which is still expanding is made up of argumentative essays written by learners within the range of higher intermediate to advanced learners of English. In its present form, its texts are produced by learners from 16 L1 background namely: Tswana (South African language), Turkish, Bulgarian, Chinese, Japanese, Norwegian, Czech, Dutch, Polish, Finish, Russian, French, Spanish, German, Swedish, German, Italian and Swedish. The LINDSEI, is also a collaborative project between several universities internationally. It is made up of over 1 million words of informal interviews transcripts produced by higher intermediate to advanced learners of English. About 80% of the texts were produced by learners, representing 11 different mother tongue backgrounds (Gilquin, De Cock & Granger, 2010). Below is a table containing some of the well-known open access learner corpora.

Table 2.1. Some of the existing popular learner corpora.

Corpus	**L1**	**Location**
International Corpus of Learner English	Various	Belgium
Louvain International Database of Spoken Interlanguage Database (LINDSEI)	Various	Belgium
The International Corpus Network of Asian Learners of English	Various	Asia
Chinese Learner English Corpus	Chinese	China
EVA Corpus	Norwegian	Norway
JPU Corpus	Hungarian	Hungary
Learner Corpora at the Language Bank	Various	Hong Kong
The Multimedia Adult English Learner Corpus	Various	USA
The Corpus Archive of Learner English in Sabah/Sarawak	Malay	Malaysia
The Montclair Electronic Language Learners' Database	Various	USA
Singapore Corpus of Research in Education	Various	Singapore
The Uppsala Student English corpus	Swedish	Sweden
The Arab Learner English Corpus (ALEC)	Arabic	Kuwait
Belarussian Learner Corpus of English	Belarusian	Belarus
The MERLIN corpus	Various	EU
The Barcelona English Language Corpus (BELC)	Spanish	Spain
Corpus of Chinese Learner English (CCLE)	Chinese	China
Taiwanese Learner Corpus of English	Mandarin	Taiwan
The Japanese Learner English Corpus	Japanese	Japan
Hong Kong Corpus of Spoken English	Chinese	Hong Kong
Trinity Lancaster Spoken Learner Corpus	Various	UK

In a nutshell, what this literature review has revealed is a gap in the literature on collocational studies from Nigeria. The few studies from Nigeria seem obsessed with collocational error analysis completely neglecting the process of acquiring collocation. There are gaps, too, in the methodologies adopted in these studies as well as the definition of the proficiency levels

their texts represent. None of these studies used computer learner corpus data and corpus analysis software in their research. The existing literature on L2 collocations globally seems to have neglected collocations in the New Englishes. The literature review has also revealed a gap in computer learner corpora research in Nigeria particularly the compilation of Nigerian learner corpus. These are some of the gaps this study attempts to fill.

Chapter Three

Pilot Study

3.0 Introduction

This chapter presents the pilot study which precedes the main study and discusses how it helps to shape the research questions and the research method applied in this study. The broader aim of my main research is to investigate the production of collocations by Yoruba-speaking Nigerian learners of English as a second language; to identify the most problematic collocations in Nigerian English language output and the causes of such problem if any. However, considering the fact that Nigeria is a big multilingual country with the existence of Nativised English and Pidgin English, the complex linguistic landscape constitutes both an opportunity and a challenge for linguistic research. In view of these complexities, a decision was made to conduct a pilot study before embarking on the main study to have a preview of Nigerians' knowledge of collocations. The intention was to have a clearer picture of the Nigerian context so as to make informed decision as to the scope and the appropriate instruments and procedures to be used in the main study.

Pilot study in Applied Linguistics as well as other related fields can be used as a "small scale version or trial run in preparation for a major study" (Polit, Beck, & Hungler, 2010: 467). It is conducted before the main research study in order to ensure that the research instruments and procedures work as they are intended. Pilot study is invaluable in determining the practicality of data collection procedures and in identifying problems before embarking on the actual study (Mackey & Gass, 2005). It can also be used to enhance the validity and reliability of the research instruments (Cohen, Manion, & Morrison, 2000). This includes checking the statistical and analytical processes to determine if they are efficacious (Simon, 2011). In addition to the above, a pilot study may also address a number of logistical issues about the research.

Generally, there are two different populations that second language acquisition researchers may draw on in a pilot study. By population, I mean the entire set of people, texts, and so forth that comprise the focus of a research study. The first group which researchers can draw on is a representative sample of the population for which the instrument is intended while the second group is a baseline group by which I mean the control group (Loewen & Plonsky, 2015). I will now present the pilot study.

3. 1 Background to the Study

While L2 collocation research elsewhere has shown that second language speakers have problems producing acceptable collocations (Bahns and Eldaw 1993; Nesselhauf, 2005), there is no comprehensive research on the collocational competence of Nigerians for whom English is the official language. This pilot study is, therefore, born out of the need to have a preview of Nigerians' knowledge of collocations, as I set out to investigate the acquisition of collocations by Yoruba-speaking Nigerian learner of English as a second language.

The population of this study comprises of two groups of Nigerian advanced speakers of English. Sixty respondents voluntarily participated in the study. Thirty of them are Nigerians who have been residing in the UK for at least three years and up to twenty years while the other thirty are Nigerians residing in Nigeria. Half of the UK group has postgraduate qualifications, while the other half has undergraduate qualifications. All the respondents (both the UK and the Nigeria groups) have a credit pass in English language in the West African Secondary School Certificate Examinations and also have a minimum of first degree with English as language of instruction.

Although the composition of the participants was essentially based on availability, educational qualifications, easy accessibility and willingness to volunteer, it was import that the population reflects the linguistic complexity of Nigeria. Hence the participants are drawn from speakers of Yoruba, Urhobo, Isoko, Edo and Igbo as their L1. These are some of the major languages of southern Nigeria. This was to ensure that the population of the pilot study was representative sample of the population for which the instrument of the main research is

intended. The choice of two groups of participants was to provide for comparison of the effect of context (immersion in the case of the UK residents) on the acquisition and production of L2 collocation. A pilot study, ideally, should be a relatively small study but the scope and depth of this study means it is a main study in some sense. The rationale for this was that a pilot study with a relatively broader scope was necessary considering the size and linguistic complexity of Nigeria to reveal all that needed to be known in order to make informed decision regarding the research design and methodology for the main study. The pilot research which focuses on Nigerian advanced speakers of English aims to answer the following research questions:

1. Do Nigerian Advanced Speakers of English have problems producing acceptable collocations?
2. Is there any correlation between the length of stay in the UK and the participants' collocational competence?
3. Which types of collocations are most problematic for Nigerian Advanced Speakers of English?

3.2 Research Method, Design and Procedures

Determining the most appropriate instrument to investigate the above research questions was the next challenge. Second language researchers have used various instruments to assess second language proficiency. One of such instruments is cloze test. Data from a wide variety of sources have supported the cloze test technique as a global measure of language proficiency (Oller, 1973). Close test can produce diagnostic information on L2 speakers' language skills. It is sufficiently integrative, and suitable not only for assessing morphosyntactic competence but also lexical and discourse competence. This makes it suitable for assessing collocational competence of L2 language speakers. Although there exists some controversies as to what aspect of linguistic competence cloze tests measure, testing research has shown that cloze tests scores tend to correlate highly with standardized proficiency scores (Bachman, 1985). Some of the issues with cloze test are essentially about the distance between blanks, scoring methods, difficulty levels, and grammatical categories of deletions (Oller, 1973). However, this cloze test was carefully designed bearing in mind these issues so that the participants could supply the words (collocates) deleted by tapping into their background schemata and making them to think critically about the missing

collocates to reconstruct the mutilated passage and, in the process, producing the collocations. Besides, it is highly adaptable to various L2 proficiency levels and contexts by manipulating the difficulty level of the test and the wording to assess specific linguistic features in this case: collocations. This makes cloze test more suitable for the pilot study.

In designing the cloze test, I selected over forty possible collocations and then used the BBI Dictionary of Word Combinations (Benson, Benson & Ilson, 1986), the British National Corpus, the Corpus of Contemporary American English (COCA) and the Corpus of Global Web-based English (GloWbE) – a 1.9 billion word corpus from 20 countries – to establish that the selected combinations were actually collocations based on their statistical frequency in the above corpora, compositionality, and the substitutionability of their constituent parts. The next thing was the creation of a database of possible collocates for each of the nodes of the selected collocations. The 40 collocations that made it to my final list were categorized into grammatical and lexical collocations, and then lexically profiled and sub-divided into: semantically opaque/semantically transparent and congruent/incongruent collocations. Grammatical collocations are defined as consisting "of a dominant word – noun, adjective/particle, verb – and a preposition or grammatical construction" (Benson, Benson, and Ilson, 1997: ix). Benson, Benson and Ilson's (1986) grammatical collocations fall into the following combinations: noun + preposition, noun + to + infinitive, noun + that – clause, preposition + noun, adjective + to + infinitive, adjective + that – clause, and the English 19 verb patterns (see BBI Combinatory Dictionary of English for more details). Examples of grammatical collocations are: *adhere to, by accident, good at, apathy towards,* etc. Lexical collocations on the other hand consist of nouns, adjectives, verbs and adverbs. Examples of lexical collocations are: *reach a verdict, commit murder, withdraw an offer, make/create an impression,* etc.

After the categorisation of the 40 collocations, I then designed the cloze task – a 40-item stimulus response collocation test. This task was a short fictitious story I composed, incorporating all the collocations in the story. The story was set in rural Nigeria, a familiar context for the participants. Since the singular purpose of the cloze test was to assess the collocational competence of the participants, it was important for the diction and the context of the test to be familiar enough. In this way, all the participants would have the schemata to

engage with the task since people are more likely to notice things that fit into their schema (Burgin, 2016). This was necessary to ensure participants' performance in the test was not negatively influenced by lexico-semantic and contextual factors that were outside the linguistic and cultural frame of reference of the participants.

All the collocates in the task were deleted leaving only the nodes, and the participants were required to provide the missing collocates. The context was explicit enough to prompt the participants to produce the acceptable collocate (s) if they knew them because the idea expressed in the story is something Nigerians are familiar with. So, the constructs were operationalized as the ability to produce single word acceptable collocate in response to the stimulus word, in this case, the node. This was to ensure the test focused only on measuring the participants' ability to produce acceptable collocations. The sixty participants would, technically, produce 2,400 collocations (40 X 60 = 2,400). To validate the test, it was administered to a native speaker of English. He was asked to provide the missing collocates within a time frame similar to the one given to the participants to complete the cloze task. Based on his feedback, the database of the possible collocates was expanded.

3.3 Test Procedure, Scoring and Analytical Approach

The test was administered to the participants with an instruction not to consult any reference materials, and to be completed within an hour. In order to get as reliable results as possible, participants were not informed the test was aimed at evaluating their knowledge of collocations until after the test. To ensure the reliability of the scoring system, uniform assessment criteria was established as follows:

- Any gap left blank was to be considered as wrong.
- Any combination which does not have any instance(s) of such co-occurrence in the British National Corpus (BNC), the Corpus of Global Web-Based English (GloWbE), the Corpus of Contemporary American English (COCA) or the BBI Dictionary was to be regarded as wrong.
- Any paraphrasing instead of producing the one word collocate was to be considered wrong even if such paraphrasing was intelligible.

- Any collocate rendered in either present or past tense was to be considered as correct even if the tense structure was wrong as long as it was the right combination.
- Any wrongly spelt collocates were to be considered as correct as long as they were the right combination.
- Any gap that was filled with a non-existent English word was to be considered wrong.

I statistically analysed the results for the whole population as well as the differences between the sub-groups. In order to gain insights to group-based performance, I ran series of Independent Samples *t-test*s to compare mean scores of the groups in relation to the variation in the data. I will now present the findings of the pilot study.

3.4 Findings of the Pilot Study

The descriptive statistics for collocation appropriacy (i.e. the score out of 40) for the whole population (n = 60) are shown in Table 3.0.

Table 3.0 Statistics for the whole Population

	Results
N	60
Number of collocations	40 (2,400)
Maximum score	40
Minimum score	14
Mean	27.03
SD	5.81

The entire population produced a total of 2,400 collocations (60 participants multiplied by 40 collocations), 778 of which were considered unacceptable while 1,622 were considered acceptable representing a mean score of 27.03 and standard deviation of 5.81.

3.4 The Nigeria Group versus the UK Group

The collocations output for the two groups were analysed in order to get an overall picture of the difference in their production of the appropriate collocate and a clearer picture of in-group performance. With each of the participants producing 40 collocations, each of the group collectively produced 1,200 collocations. Out of this number, the Nigeria group produced 864 acceptable collocations representing a mean score of 28.80 and standard deviation of 5.63. The highest score in the group is 40 and the lowest is 18. Eleven out of the thirty participants representing 36% of the group scored over 31 with four of them scoring over 35. Ten participants in the group representing 33.3% scored between 26 and 30. The UK group, on the other hand, produced 758 acceptable collocations representing a mean score of 25.26 and standard deviation of 5.52. The highest score in the group is 34 and the lowest is 14. In sharp contrast to the Nigeria group, only five participants representing 16% scored above 31 with only one scoring above 35. Nine participants in the group representing 30% scored between 26 and 30.

While the Nigeria group produced 336 unacceptable combinations, the UK group on the other hand produced 442 unacceptable combinations. That is 106 more non-acceptable collocations than the Nigeria group. Taking 26 out of 40 as the cut off mark considering their general English proficiency level, 33.3% of the Nigeria group and over 50% of the UK group have collocational deficiency. An independent sample *t-test* confirmed a statistically significant difference between the two groups for acceptable collocations production at $t\ (58) = 2.452,\ p = 0.017$. See Table 3.1 for group statistics bellow for more details.

Table 3.1 Pilot Study Group Statistics on Collocation Production

Group Statistics					
	Grouping	N	Mean	Std. Deviation	Std. Error Mean
Collocations	Nigeria group	30	28.80	5.635	1.029
	UK group	30	25.27	5.527	1.009

3.5 The Production of Lexical Collocations

There are twenty-six lexical collocations in the cloze task. So, with each participant producing 26 lexical collocations, the two groups collectively produced 780 collocations each. The Nigeria group produced 487 acceptable collocations representing a mean score of 16.23 while the UK group produced 411 representing a mean score of 13.70. The highest score in the Nigeria group is 26 and the lowest is 6. The UK group has 25 as the highest score and 6 as the lowest. Of the 780 collocations produced by each group, the Nigeria group produced 293 unacceptable collocations while the UK group on the other hand produced 369 unacceptable collocations. That is 76 more non-acceptable collocations than the other group. An independent samples *t-test* indicates there is no statistically significant difference between the two groups for acceptable lexical collocations production at t (58) = 1.952, p = 0.056.

3.6 The Production of Grammatical Collocations

The same analysis was carried out on the Grammatical Collocations sub-group. There are 14 grammatical collocations in this group. Both groups did very well in this category. Collectively, each group produced 420 grammatical collocations. The Nigeria group produced 374 acceptable collocations representing a mean score of 12.46 which means 89% of the grammatical collocations produced are acceptable. Only 46 of the grammatical collocations produced by this group are considered unacceptable. The UK group also produced good numbers of acceptable grammatical collocations. Three hundred and fifty grammatical collocations produced by this group representing a mean score of 11.66 are considered acceptable. That is 83% acceptable grammatical collocations. This group collectively produced only 70 unacceptable grammatical collocations. As we can obviously see, an independent sample *t-test* shows no statistically significant difference between the two groups for acceptable grammatical collocations production at t (58) = 1.523, p = 0.133.

3.7 The Production of Incongruent Collocations

A total of 20 incongruent collocations are involved in this study. These are collocations that have no equivalent L1 construction. The Nigeria group collectively produced 600 incongruent collocations, out of which 350 representing a mean score of 11.6 and 58.3% of the incongruent collocations are acceptable. That means 250 representing 41.7% of the 600 incongruent collocations produced are deviant. A fine-grain analysis of the in-group performance reveals that, of the thirty participants, nine of them representing 27.9% scored over 15 out of 20, collectively produced 150 incongruent collocations. Eleven out of the thirty scored less than 10 out of 20, collectively producing only 83 acceptable incongruent collocations. That suggests 36.3% of the Nigeria group have below average knowledge of incongruent collocations. The other 34% who scored between 11 and 14 could be regarded as having average knowledge of incongruent collocations.

The UK group, on the other hand, also produced 600 incongruent collocations out of which 272 representing a mean score of 9.06 and 45.4% of the incongruent collocations are acceptable. A total of 328 representing 54.6% of the 600 produced are deviant. An in-depth analysis of the in-group performance reveals that only two participants representing 6.6% scored above 15 out of 20. Seventeen of them representing 56.1% scored less than 10 out of 20. A further analysis of the UK group reveals that half of the population who have postgraduate qualifications scored more than those with undergraduate qualifications.

The summary of the statistics of the groups is as follows: 29.7% of the Nigeria group and 6.6% of the UK group have good knowledge of incongruent collocations; 34% of the Nigeria group and 37.3% of the UK group are just within average; and 36.3% of the Nigeria group and 56.1% of the UK group have below average knowledge of incongruent collocations. Looking at the entire population, based on this data, we may say 18.15% of the participants have good knowledge of collocations, in terms of their incongruent collocational competence, 35.65% are average, and 46.2% lacks collocational competence. See Table 3.2 for the group statistics for more details.

Table 3.2 Pilot Study Group Statistics on the Production of Incongruent Collocations

Group Statistics					
	Grouping	N	Mean	Std. Deviation	Std. Error Mean
Incongruent Collocations	Nigeria Group	30	11.67	4.080	.745
	UK Group	30	9.07	3.413	.623

An independent samples *t-test* confirmed a statistically significant difference between the two groups for incongruent collocations production at t (58) = 1.677, p = 0.010.

3.8 Discussion and the Implications of the Findings

In this section, I discuss the findings of the pilot study in the light of the existing literature on L2 collocations research. The discussion will focus on the implications of these findings for my main study and how they helped to shape the scope, the research questions, the research design and the instruments used in my main study, as well as the choice of population for the research.

I set out to explore L2 collocational competence of Nigerian advanced speakers of English as a second language in a way that is not limited to error analysis. My aim was to use an elicitation measure that would enable the collection of 2,400 collocations from two groups of Nigerians (UK group and Nigeria group). I wanted to investigate the correlation between their length of stay in the target language environment and their productive knowledge of collocations. I also wanted to know if users of English as a second language (Nigeria's official language) as opposed to English as a foreign language would also have difficulty producing acceptable collocations. If confirmed to be so, I wanted to get some insight to the types of collocations Nigerian advanced speakers would find more problematic.

My results show that of the 2,400 collocations collectively produced by the entire population, 32% of them were unacceptable. Considering their general English proficiency level, one would expect a higher score. This could be indicative of problems in producing acceptable collocations despite being advanced speakers of English. They are regarded as advanced speakers because they all have at least an undergraduate university degree taken through the medium of English. Another interesting thing here is the fact that none of the words constituting of the collocations tested in this study would be strange to the participants, considering their educational qualifications and experience, goes on to confirm that L2 collocations knowledge lags behind their general vocabulary knowledge.

While the performance of sixty Nigerian advanced speakers of English as a second language may not be enough to make some big claims about the collocational competence of Nigerians as a whole, nonetheless, the findings tend to suggest that collocations might be a problematic phenomenon for Nigerians. This is more pronounced in their production of incongruent collocations. Based on these findings, it may be plausible to conclude that collocations which have no equivalent L1 construction are most problematic for Nigerians. This is consistent with various findings in the literature (Bahns, 1993; Farghal & Obiedat, 1995; Yamashita & Jiang, 2010).

Collocations are prevalent in language, particularly the English language, and the fact that we hear them and even produce them every day should have made them easy to acquire but on the contrary, they have become one of the most problematic linguistic phenomena for second language users. The difficulty in handling them seems to stem from their ambiguous linguistic status and the lack of rules governing their formation. Even highly proficient second language users resort to direct L1 to L2 translation when producing incongruent collocations in some cases as shown below:

"Right from my first night in the hostel, I started 'writing a diary' ….**; …** I started 'jotting a diary'….; I started 'making a diary'".

The collocation: *keeping a diary*, has no equivalent construction in the participants' L1, what we have is literarily 'writing a diary'*;* and 22 participants produced 'writing a diary' while nine of them produced 'making a diary' or 'jotting a diary'. Although the other 29

participants produced the acceptable construction – *'keeping a diary',* the statistics is a confirmation of the strength of the "gravitational pull of the mother tongue" (Salim, 2001: 117) in the production of L2 collocations.

Whenever the participants could not interpret or chose not to use direct L1 to L2 interpretation in their production of incongruent collocations, they resorted to overgeneralization as shown in the extract below:

"He would tell a story of a 'fleet of lions'; 'a pack of lions' that used to roam the forest before uncontrolled timber cutting destroyed their habitat. He told of a time when the villagers mistook a 'group of whales';' 'a mob of whales' that often came close to the shore for the colonial masters' submarines".

The words: *fleet, pack, group,* and *mob* all refer to large number of things, but it is not acceptable to use them as seen in the extract above. We may say a *fleet of cars* or *pack of dogs*, but it is unnatural to say a 'fleet of lions' or 'pack of lions'. We may say a *pod of whales* or a *school of whales,* but natives or proficient speakers of English may not say mob of whales.

Other non-acceptable collocations which are pervasive in the participants' output are shown below:

"This was in 'opposite contrast' to my neighbour's parent".

"… to deter other from 'contradicting the law'".

"… any object with which we could 'sustain/incur a wound' on ourselves"

"He would ask us a few questions to 'gain our attention' to the morals of the story"

"The high court 'annulled his appeal'".

"My father's stories 'doused/minimized my fear' and prepared my mind to 'adapt with' this change.

"Recite music', 'unease my fear', 'adverse poverty', 'disallowed his appeal', 'proof our attention', 'ascertain attention'"

'Discarded /annulled his appeal'.

"She ran into the burning house with 'haste/ absolute/resolute abandon' to rescue her youngest child"

'Reckless abandon' is not a universally frequently used collocation according to frequency data from the Global Web-Based Corpus of English (GloWbE). However, it is frequently used in Nigeria more than any English-speaking countries in the world. One would expect this would mean the participants should be able to produce it correctly. But out of the sixty participants, only nineteen produced the acceptable construct. It is not clear why this is the case. The construct: 'abject poverty', on the other hand, is also a frequently used collocation in the GloWbE, most frequently used in Nigeria, Ghana and Tanzania. Unlike 'reckless abandon', almost all the participants produced 'abject poverty' correctly. Is this an indication that high frequency of occurrence is a fairly reliable predictor of mental representation? Is this an indication that repeated exposure to collocations may enhance acquisition as evidenced in the case of 'abject poverty'? Though only few of the participants were able to produce 'reckless abandon' correctly despite it being a frequent expression in Nigeria, there is slight evidence in this study to suggest that the participants have less problem with frequent collocations than the infrequent ones. This seems to indicate that the more they are exposed to the constructs the more they are likely to recall it while producing the language. But this will be an area that will be investigated further in the main study.

It is traditionally assumed that the best way, if not the only way, to acquire second language collocations is to spend an extended period of time in the target language environment working or studying. This, it is believed, would facilitate maximum exposure to the target structure which would consequently translate to acquisition. However, this immersion-based approach to the acquisition of L2 collocations has been called to question by the publication of large scale studies like the one carried out by Nesselhauf (2005). Her corpus-based study of collocation usage among advanced-level German EFL students shows that 'increased exposure to English in English-speaking countries leads to a slight improvement' and 'the length of stays in English-speaking countries does not seem to lead to an increased use of collocations; instead, there even seems to be a slight trend in the opposite direction' (Nesselhauf, 2005: 236). This study seems to corroborate her findings. The UK group in this study appears to lag behind the Nigeria group in their productive knowledge of collocations.

However, the scope and depth of this study is not enough to conclude that Nigerian advanced speakers of English as a second language who reside in English as L1 context are more proficient than Nigerians living in Nigeria where English is L2. But the study confirms that L2 collocations are problematic for second language users, and incongruent collocations are more problematic. Living in an English-speaking country does not necessarily facilitate speedy acquisition of collocations as these findings suggest.

While semantically opaque collocations are believed to be problematic for learners, this study does not reveal any convincing evidence to confirm that is the case with the population. The reason for this could be that advanced speakers have sufficient semantic knowledge of the lexicon of the English language. In a nutshell, this pilot study has revealed evidence of considerable collocational deficiency among Nigerian advanced speakers of English particularly in incongruent collocations. It also confirms collocational knowledge is not easily acquired even when living in the target language context as an adult second language speaker. What then are the implications of these findings?

These findings have revealed many factors that need to be considered while designing the main study. One factor could be that because the participants in the pilot study speak five different L1, it is difficult to determine which of the L1s is responsible for any of the L1 influenced errors in the data, hence making it difficult to analyse the data to determine whether the collocations have L1 equivalent or not. In view of this realisation, the best approach to my main study would be to get data from a population that speaks the same L1. The second factor is that though cloze test is a good instrument for assessing L2 proficiency, it however does not allow for participants to freely produce the language at a scale that will truly reflect the extent of their collocation knowledge. The way around this is by using corpus data. Learner corpus is a versatile new source of data for second language acquisition research (Granger, 1998). While designing and building a corpus, it is possible to control the age, the mother tongue, the context of learning and other variables of the participants. A learner corpus based on clear design criteria lends itself particularly to a contrastive analysis (CA) – not traditional CA but Contrastive Interlanguage Analysis (Granger, 1996; Granger, 2015). This concept will be discussed in-depth in the section on research design and procedure. In essence, a corpus-based method provides for a wider scope. However, there is

no Nigerian corpus of English produced by homogenous population. The Nigerian component of the Louvain based International Corpus of Learner English (ICLE) and Corpus of Global Web-based English (GloWbE) is made up of text produced by Nigerian speakers of various L1. So, they would not be suitable for this research, hence the need to build a new corpus.

The pilot study also reveals Nigerians have some difficulty producing acceptable collocations and as such, the main study will continue that strand of investigation. However, although the study did show that Nigerians who are apparently immersed in the L2 context (the UK group) are less proficient in L2 collocation, the main study will not investigate the reason for the disparity in the collocation proficiency of the two groups. Pursuing this strand of investigation may change the focus of the main study and widen the scope beyond what is intended.

Investigating the collocation competence of some randomly selected Nigerian advanced speakers of English as it was done in the pilot study means it is not possible to look at the developmental selectivity of LI influence on L2 acquisition as it manifests at various proficiency levels. The principle of selectivity refers to "formal properties that make L2 structures immune or receptive to L1 influence as well as L2 developmental stages that activate L1 transfer along a time axis" (Zobl, 1980a: 43). This concept emanates from the structuralist assumption that a language will accept only those external influences that correspond to its own structural tendencies and systemic biases. According to Zobl (1980b: 469) "when an L1 structure conforms more closely to general acquisition regularities or processes than the L2 structure to be acquired, then the occurrence of transfer is promoted". Through Contrastive Interlanguage Analysis, the areas where an L2 is potentially susceptible to L1 influence can be identified. This is all about identifying and analysing L1 influence on the production of L2 collocations. In view of the above, the main study will not use randomly selected Nigerians but rather select a research population that will represent various proficiency levels in order to thoroughly investigate the developmental selectivity of LI influence on L2 collocation acquisition. And a close test will not be used as the research instrument in the main study. Having presented the pilot study and discussed how it helped to shape the main study, the next section will now focus on providing overview of the various

research methods that have been used to investigate collocation. It will also discuss my research design and methodology, the procedures, the justification for using the methods as well as how each part of the methods addressed my research questions.

Chapter Four

Research Design and Methodology

4.0 Introduction

This chapter has four main sections. The first section presents the research questions which this study is investigating. The second section provides an overview of the various methods that have been used in L2 collocation research, focusing on corpus-based method and providing the justification for using corpus-based approach in this study. The study corpus – the Nigerian Learner Corpus of English (NILECORP) – is also presented in the second section. It explains the design criteria and procedures for building the learner corpus from ethics approval to defining and describing the population as well as data elicitation, data capture and text handling, converting written materials into electronic format and the assignment of proficiency levels to the corpus texts. The third section also focuses on the reference corpora – The Louvain Corpus of Native English Essays (LOCNESS) which is the primary reference corpus and the Nigerian component of the Corpus of Global Web-based English (Davies, 2013) – the secondary reference corpus. The third section ends with the justification for using these corpora as the reference corpora for this study. Finally, the fourth section describes the approaches and procedures used for the extraction of collocational candidates from the study corpus and the primary reference corpus as well as the analytical approaches used for analysing the data and how each aspect of the research method addresses my research questions.

4.1 Research Questions

1 To what extent do native and non-native writers make use of collocations?

   (a) Do native speakers use more collocations than L2 learners?
   (b) What is the difference between the collocations produced by the two groups in terms of the linguistic complexity of their constituents?

(c) How many semantically burdensome collocations are produced by both groups? These are collocations with a range of connotative and associative meanings; their meanings have been modified to introduce additional nuances and associations.
(d) How many congruent and incongruent collocations are produced by the L2 learners?

There will be a detailed comparative analysis of all the Verb Noun and Adjective Noun collocations produced in the LOCNESS and the most proficient of the four learner groups (this learner group, as we will find out later in this study, is equivalent to the Common European Framework of Reference for Languages proficiency level C1).

2. Is there a relationship between frequency of and exposure to input in L2 learners' speech community and their production of collocations?
   (a) What effect does the frequency of input in the learners' speech community have on their production of verb noun collocations?
   (b) What effect does frequency of input in the learners' speech community have on their production of adjective noun collocations?

This will further elaborate on the findings of the first sets of questions above considering how frequency of input affects the collocational production of L1 Yoruba learners of English.

3. What is the relationship between proficiency and the production of collocations?
   (a) What is the relationship between proficiency and the production of verb noun and adjective noun collocations?
   (b) What is the relationship between proficiency and the production of incongruent verb noun and adjective noun collocations?
   (c) What is the relationship between proficiency and the production of congruent verb noun and adjective noun collocations?
   (d) What is the relationship between proficiency and the use of linguistically complex verb noun collocations in terms of the collocation span and the structural properties of their constituents?
   (e) What is the relationship between proficiency and the use collocations with additional nuances and associations – the degree of semantic opacity and transparency?

As part of this investigation, I will analyse the verb noun and adjective noun collocations produced by four groups of Yoruba-speaking English learners representing four proficiency levels.

4. What is the nature and causes of the errors in the collocations produced by the learners?
   (a) What types of collocations are the most problematic for the Learners?
   (b) What is the nature and causes of the collocational errors in the Learner Corpus?
   (c) What are the similarities and differences in the error across proficiency levels?
   (d) What proportion of collocation errors are due to: [a] Inter-lingual factors and [b] Intra-lingual factors.

The focus of these questions is on the identification, classification and analysis of all the erroneous verb noun and adjective noun collocations extracted from the learner corpus.

The collocations which would be investigated in this study are based on Benson, Benson, & Ilson's (1986) categorisation. In order to make the scale of the study manageable, I will focus on two types of lexical collocations namely: verb noun and adjective noun collocations. I will now discuss the systematic approaches and the instruments used for investigating these questions.

4.2 Collocations Research Methods

Three main methodologies have been used in the literature to investigate the knowledge and use of collocations by L2 learners. They are the elicitation of collocations through various elicitation techniques (Bahns & Eldaw, 1993; Farghal & Obiedat, 1995); the traditional error analysis of samples of learners' output (Ridha & Al-Riyahi, 2011; Yumanee & Phoocharoensil, 2013; Ha, 2013); and the analysis of learner corpora using various corpus analysis tools (Grainger, 1998; Nesselhauf, 2003, 2005; Schmitt & Siyanova, 2008; Groom, 2009; Laufer & Waldman, 2011). More than half of published studies on collocation in

learner language are based on elicitation tests and mainly on productive data (Nesselhauf, 2005).

Elicitation studies of collocation focusing on the question of what L2 learners can produce have used either cloze tests or translation tests or a combination of both techniques (Biskup, 1990, 1992; Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Herbst, 1996; Shei, 1999). These elicitation techniques include: multiple choice tests, gap-fill tests, appropriateness judgment tests and translation tests. Some of the studies that used gap-fill tests provided the first phoneme of the collocate and required the learners to complete the missing part (Al-Zahrani, 1998), while others provided the node and required the participants to provide the missing collocate (Bahns & Eldaw, 1993). The rationale behind the method is that, if the subjects knew the collocations, the provision of the first phoneme of the collocate or the node should prompt them to produce the collocation. Meanwhile, other studies have used a combination of cloze test and translation tests (Farghal & Obiedat, 1995). Biskup (1992), for instance, asked 34 Polish and 24 German Advanced learners of English to translate 23 collocations from their L1 to English. While many researchers have used these elicitation techniques, the main limitation is that it is only suitable for small data. This probably explains why most of the elicitation studies of collocations have concentrated on few sub-sets of collocations (Biskup, 1990, 1992; Shei, 1999). Besides, these studies have not analysed their results in more detail apart from Farghal and Obiedat (1995) who analysed the collocations produced by Advanced Arabic-speaking learners of English.

Some other studies relied on the well-used approach of traditional error analysis of samples of learners' output. Collocational studies that used this method identified and isolated deviant word combinations which they regarded as errors and analysed them to determine the causes of the errors (mainly L1 interference). Most of the collocational studies from Nigeria used error analysis method (Taiwo, 2001; Okoro, 2013). An error can be defined as a deviation from the norms of the target language (Ellis, 1994). As simple as this may seem, this definition is, however, problematic in the sense that it raises a number of questions. First, there is the question regarding which of the existing varieties of the target language should serve as the norm. Should it be one of the prestigious varieties (British or American English) or one of the emerging Englishes? In the Nigerian context, for instance, one will also have to

consider the varieties the learners are exposed to as well as the sociolinguistic reality of language use in Nigeria. The second question concerns the distinction between errors and mistakes. According to Ellis (1994), an error takes place when the deviation arises as a result of lack of knowledge; it is a reflection of lack of competence. A mistake on the other hand, occurs when learners fail to perform their competence. This means, a mistake is a performance phenomenon; it is a processing failure. It could be due to memory limitations and lack of automaticity (Ellis, 1994). This means, learners may make mistakes in their language production and that does not necessarily translate to lack of proficiency in the language. Brown (2000: 217) states that "mistakes must be carefully distinguished from errors of a second language learner". This is very import to ensure the validity of the results of collocational studies that are based on error analysis methods. This obviously lays a heavy burden on researchers to discern between what is an error and what is a mistake. In this study, any instance of consistent deviation from acceptable norms in the Nigerian context of language use will be regarded as error.

The third method which is widely used in the literature to investigate L2 learners' collocational competence and development is learner corpus-based method. Learner Corpus Research uses the main principles, tools and methods from corpus linguistics to provide improved description of learner language which can be used for second language acquisition research and language teaching (Granger, 2002). Granger (2002) defines corpus linguistics as a linguistic method which is founded on the use of electronic collection of naturally occurring texts. This collection of samples of naturally occurring language (texts of written and/or spoken language) presented in electronic form is known as a Corpus (Hunston, 2006). Learner corpus, therefore, is a "systematic computerized collections of texts produced by language learners" (Nesselhauf, 2005: 40). Being systematic means the texts that constitute a learner corpus are selected based on certain criteria often determined by the aim of the study for which the corpus is compiled.

Corpus research has led to a much better description of many of the different registers as well as various dialects of native English (Granger, 2004). Leech (1992: 106) describes corpus research as a "new research enterprise, [. . .] a new philosophical approach to the subject [. . .] an 'open sesame' to a new way of thinking about language". This is more so because of

the power of computer software tools combined with the impressive amount and diversity of naturally occurring language data used as evidence which has revealed many linguistic phenomena which are hitherto unknown. Stubbs (1996: 232) attests to "the heuristic power of corpus methods" which "have led to far-reaching new hypothesis about language, for example about the co-selection of lexis and syntax".

There is a consensus in the literature today that corpus data is the most reliable source of evidence for such features as frequency (McEnery & Wilson, 1996). The strength of corpus-based methodology lies in its suitability for conducting quantitative analyses (Granger, 2004). Methodologies commonly associated with learner corpus research are the comparison of native and second language learners of a language, and different types of L2 learners of a language. With various corpus analysis tools, researchers can search a corpus or (or corpora in a comparative study) "for a given target item, count the number of instances of the target item in the corpus and calculate relative frequencies, display instances of the target item so that the corpus user can carry out further investigation" (Hunston, 2006: 234).

4.2.1 The Justification for a Corpus-based Method

Having examined the other major methods that have been used in the literature to investigate L2 collocations, the most suitable method for achieving the overall aims of this research is a corpus-based method. One of the strongest justifications for this method is the fact that learner corpus is a very rich type of resource which lends itself to a wide range of analyses. It integrates both qualitative and quantitative analytical techniques. This is very important considering the scope and nature of this study, and as such the resource that lends itself to various linguistic analyses can best answer the research questions.

Moreover, learner corpus data is more reliable in the sense that it pools together the linguistic intuitions of a range of L2 speakers thereby offsetting the potential biases in the intuition of an individual speaker. It is more natural since corpus data is language used in real communication and not invented for specific linguistic analysis. In comparison to intuition,

corpus data can find differences which intuition alone cannot perceive. The corpus-based approach, by nature, is empirical, analysing the actual patterns of language use from natural texts. All these make corpus-based method the preferred method for collocational research.

The study reported in this thesis is based on a wide empirical base focusing on the language of a large numbers of participants from two contexts – English as a native language and English as a Second Language. A population that is large enough and representative of the speech community is necessary to generalize the results. The empirical nature of corpus data makes it ideal for this purpose. It pools together the intuitions of a large population of speakers and makes linguistic analysis more objective (McEnery & Wilson, 2001; McEnery & Xiao, 2011). This fits in with the objectives of this study. Other elicitation techniques such as cloze tests or translation tests would not have produced the sort of resource described above. Data sources such as experimental or introspective data would not be samples of natural language use. Besides, it would be practically impossible to get experimental or introspective data in a study which focuses on the language output of a large population.

4.3 The Study Corpus

4.3.1 Defining the Sample of the Population of the Corpus

Defining the sample is very important when assembling a learner corpus. Nigerian secondary school students in state schools are the target sample of this research. There are over 7,000 public secondary schools in Nigeria spanned across 36 states and the Federal Capital Territory (Abuja) with over 3.2 million students. This, obviously, is a very large population and studying the whole population is impracticable. This is where sampling comes in. Even then, this is not a straightforward process considering the linguistic complexity of Nigeria where the people speak over 521 languages. This means having corpus data produced by a sample that is heterogeneous may be difficult to analyse. This is because their various L1s may influence their L2 production (L1 interference) and it will be practically impossible to know which of the L1s is responsible for any deviation in the language output. Because of this, a sampling frame was drawn up to guide the application of sampling to select a

manageable and representative subset of the target population. The sampling frame contains all the elements the population of interest must have. These elements are: participants must be from state schools, must be speakers of Yoruba language as mother tongue or with native-like proficiency and must be in certain language proficiency levels. The above criteria for selecting the sample are necessary to ensure all the participants are as homogeneous as possible and have similar exposure to the target language. With the sampling frame clearly defined, a stratified random sampling was then used to select 26 secondary schools from the city of Lagos in Nigeria. The strata were formed based on the participants' shared attributes as defined in the sampling frame. Ideally, a random sample from each stratum was to be taken in a number proportional to the stratum's size when compared to the population but in this case, a stratified random sampling could not be applied in its strict sense due to lack of reliable statistics on the number of state secondary schools in Lagos city as at the time of gathering these data. Selecting 26 schools in the city of Lagos out of about 319 state secondary schools in Lagos state (comprising Greater Lagos and other cities) might be arbitrary, but the overarching consideration is to have a population that is as representative as possible. And the 26 secondary schools represented over 2,000 participants. This is sufficient considering the scope of this study.

In conclusion, since the participants either speak Yoruba as their native language or their dominant language, it is more appropriate to define the sample of this study as Yoruba speaking Nigerian state secondary school students in Lagos who have been learning English in a formal setting for between 7 – 11 years. In the next section, I will describe the sample in more detail.

4.3.2 Describing the Population of the Corpus

According to Nigeria's National Policy on Education (2004), basic education shall be of 9 years duration comprising 6 years of primary education and 3 years of Junior Secondary education. These 9 years of basic education is tuition free, universal and compulsory for every Nigerian child. Upon successful completion of the basic education, students can

proceed to senior secondary education. Basically, Nigeria operates a 6 – 3 – 3 – 4 educational system. This means 6 years in primary school, 3 years in junior secondary school, 3 years in senior secondary school and 4 years in university. Children usually start their primary education at the age of 6 and proceed to secondary school at the age of 12. The primary education is for children aged 6 – 12 while secondary education is for children aged 12 – 17.

The medium of instruction in the primary school is the language of the immediate environment for the first three years (primary 1 to 3). The language of the immediate environment is the local language of the speech community. This means Yoruba language for Yoruba-speaking part of Nigeria, Hausa language for Hausa-speaking part of Nigeria, and many other local languages depending on which part of Nigeria the child lives and studies. During this period, English is taught as a subject usually for about 70 minutes a day – two sessions of 35 minutes each. From the fourth year (Primary 4 onward), English is progressively used as a medium of instruction while the language of the environment is taught as a subject. The participants in this study, therefore, have been learning the English language formally from the age of nine. However, it is important to note that because English is a second language in Nigeria, some of the participants were exposed to the language much earlier in varying degrees depending on where they live and their family's social status. Urban children are usually exposed to the language much earlier through various means such as the media, listening to interactions on the high streets, or even at home from their educated parents and siblings. Children who attended private primary schools were exposed to the language even while still in Nursery schools (pre-primary schools). However, this is not the case for rural dwellers. The rural environment with few English speakers means children in these areas have very little chance of being exposed to the language in a way that will significantly affect the acquisition of the language until they are formally taught in the primary school.

In Nigerian schools, advancement from one class to another is based on continuous assessment, and learners must pass the required assessments before being promoted to the next class. This standardised way of assessing the learners before promoting them means learners in the same class may be at the same proficiency level. This will be elaborated

further on the section on the proficiency levels of the participants considering the fact that proficiency level is a fuzzy variable in computer learner corpora (Carlsen, 2012). Meanwhile, four groups of students participated in this study. They are students in year two and three (JSS 2 & JSS 3) of Nigerian Junior Secondary Schools and students in year one and two (SS 1 & SS 2) of the Senior Secondary School.

4.3.3 Procedures for Compiling the Study Corpus

This section details the procedures for compiling the Nigerian Learner Corpus of English (NILECORP). It provides the rationale behind the various decisions that were made in the process of the compilation.

4.3.3.1 Permission: Ethics Approval and Participants' Consent

There are genuine and serious ethical issues in this study because the participants are minors. In compliance with the research ethics requirements of the university, ethical approval was sought and received from Kingston University Graduate Research School before embarking on this research. For ethical reasons, there was no direct contact between the participants and the researcher. The English language teachers in the participating schools served as intermediaries. A letter of consent was duly signed by each teacher on behalf of their students securing privacy, freedom from coercion for the participants, the teachers and their institutions, and the right to withdraw from the study whenever they deemed it necessary. In view of the above, no names of persons or institutions will be mentioned throughout this thesis

4.3.3.2 Data Elicitation, Data Capture and Data Handling

Working through the teachers, each participant was asked to write two essays of about 400 words each. I carefully chose the topics of the essays considering the learners'

sociolinguistic context and frame of reference. Some of the topics were on real life experience while others were on hypothetical instances. The themes of the essays were all familiar to Nigerian students so that they would not require additional schematic knowledge to engage with the task. This was necessary to avoid anything that might inhibit their writing ability so that they could write freely as much as possible. The following eight essay questions in two sets were given to the students with an instruction to write over 400 words from home:

1. If I had 100 Million Naira (What would I do with it?) – JSS 2
2. The day I will never forget – JSS3.
3. A friend in need is a friend indeed (a story about good friendship) – SS1.
4. If I were President of Nigeria (what would I do) – SS2.

1. My last holiday – JSS2.
2. Free Education for all: Is it a good thing? – JSS3
3. A doctor and a teacher: Which one is more beneficial to humanity? – SS1.
4. Write an essay that would end with "I wish I had listened…" (SS2).

The students had up to five days to submit their essays. They were not informed that their essays would be used for research investigating their productive knowledge of collocations. This was to avoid a situation where they would be careful while writing multiword units. This might result in them consulting reference materials while writing collocations or avoid multi-word units altogether in their essays. This obviously would defeat the purpose of the research.

Meanwhile, all the essays were submitted to their teachers who collated them and handed them over. All the scripts from the 26 participating schools were then collated into four groups namely: JS 2, JS 3, SS 1 and SS 2 representing four proficiency levels. The texts were carefully labelled to avoid mixing them up.

4.3.3.3 Converting Written Materials into Electronic Format

The texts were word processed without correcting any errors in the learners' essays. Each essay was typed out just the way it had been written. This was necessary to retain the originality of the texts. Two Word Processing companies known as 'Business Centre' in Nigeria handled the typing of the scripts because of the volume of the texts. A 'Business Centre' in Nigeria means a small shop, usually owned by one person, rendering such services as typesetting, photocopying, scanning, laminating, printing, etc. Although the people who were typing the scripts were given clear instructions on what to do, they were still monitored throughout to ensure compliance. As any mistake, such as mixing up the scripts would jeopardize the findings of this study.

Upon completion, I formatted and structured the texts (516, 917 words) to bring the data into line with corpus convention for encoding. The Learner Corpus that was built from the data was non-annotated. Corpus annotation is the addition of interpretative linguistic information to a corpus. The corpus was made up of four sub-corpora representing four groups of learners. Meanwhile, the handwritten scripts were securely destroyed in Nigeria in compliance with the terms of the ethics approval provided by Kingston University London.

4.4 Assignment of Proficiency Level to the Corpus Texts

A learner-centred method was used to assign proficiency level to the corpus texts. The learner-centred method uses the learners' characteristics and not the linguistic quality of their texts to assign proficiency levels to corpus texts (Carlsen, 2012). One aspect of Computer Learner Corpus Research which this study has attempted to do differently is the assignment and definition of proficiency levels. As I will later point out in this thesis, proficiency level is a "fuzzy variable" in learner corpus research (Carlsen, 2012: 161). Crudely labelling the groups of learners in this study as 'intermediate' or 'advanced', or 'third and fourth year high school students in Lagos' will mean little, if anything, in developmental terms and may not be interpretable in any meaningful way. In order to ensure the proficiency levels assigned to each of the four groups involved in this study is as clearly defined as possible, it is

benchmarked on the Common European Framework for Language Reference (CEFR). The CEFR was put together by the Council of Europe to describe achievements of learners of foreign language across the Europe. The Common European Framework divides language learners into three divisions which are further divided into six levels: A1, A2, B1, B2, C1 and C2. Each of these levels describes what a learner is supposed to be able to do in the four language skills of reading, listening, speaking and writing.

The table below (Figure 4.1) describes the language ability of the six CEFR proficiency levels, from the least proficient to the most proficient – four of which apply to this study. The descriptions for these proficiency levels are reproduced from the "Common European Framework of Reference for Language: learning, teaching, assessment" (Council of Europe, 2001: 24). It shows, in brief, the linguistic ability of the learners in these proficiency levels to use the language to make and communicate meaning.

Figure 4. 1 Common European Framework of Reference for Languages

A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce themselves and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.
A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics, which are familiar, or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint

	on a topical issue giving the advantages and disadvantages of various options.
C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express themselves fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express themselves spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

[Reproduced from CoE (2001: 24). Common European Framework of Reference for Languages: learning, teaching, assessment].

To determine the proficiency levels of the participants within the European Framework, copies of a self-assessment grid (see Appendix B) which illustrates the levels of proficiency described in the CEFR were sent to 39 English language teachers across 26 secondary schools in Lagos, Nigeria. They were asked to carefully read the language descriptors for the six proficiency levels in the CEFR and select the level that best describe the learners – to situate their language ability within the Common European Framework of Reference for Languages. Meanwhile, all the teachers have taught across the four groups of learners at their various schools which means they were very familiar with the language performance of the learners at these levels. All the 39 questionnaires were completed and returned but only 24 were correctly completed rendering the other 15 questionnaires invalid.

For the Junior Secondary School 2 students, there were 24 valid entries. 16 teachers placed them on A2, three placed them on A1, another three placed them on B1 while the other two teachers placed them on B2. For the Junior Secondary School 3 students, out of the 24 valid responses, 13 teachers placed them on B1. Two teachers rated them as A1, four rated them as A2, another four teachers rated them as B2 and one teacher rated them as C1. 24 valid responses were received the Senior Secondary School 1 group. 15 teachers put them at B2 proficiency level, 3 rated them as B1, another 2 rated them as C1, two teachers put them at

C2 while one put them at A2. Finally, there were also 24 valid entries for the Senior Secondary School 2 group. 14 teachers rated them as C1, one teacher rated them as A2, two teachers rated them as B1, four teachers put them at B2 and the other three teachers rated them as C2. Table 4.1 below shows the teachers' rating.

Table 4.1 Teachers' assignment of proficiency levels

Classes	A1	A2	B1	B2	C1	C2
JSS 2	3	16	3	2	0	0
JSS 3	2	4	13	4	1	0
SS 1	0	1	3	15	2	2
SS 2	0	1	2	4	14	3

4.4.2 The Assessors: English Language Teachers in Lagos

All the teachers who assessed the learners are experienced teachers who have taught across the four groups of participants. These teachers are responsible for setting the learners' examination questions, marking and grading them. So, their judgement is based on the learners' written works (examination scripts, class activities and homework), their spoken English, their reading skills as well as their listening skills. The teachers have a good understanding of their language ability. But despite this, the teachers' assignment of proficiency level to the learners is not unanimous. This could be attributed to the seemingly imprecise description of the CEFR language descriptors for the various proficiency levels. Beyond that, the divergent views of these teachers on the language ability of these groups of learners highlight how difficult it is to ascribe proficiency level to a whole group of learners. It could be easier if proficiency level was to be assigned to individual learners in each group. There may be individual variation in the proficiency level of students within each group. In order to determine if the assessors' data on the assignment of proficiency levels are

statistically significant, a chi-square goodness-of-fit test was conducted. See table 4. 2 below:

Table 4. 2 Result of chi-square goodness-of-fit test

Test Statistics				
	JSS2	JSS3	SS1	SS2
Chi-Square	22.333 [a]	18.917 [b]	29.826 [c]	23.083 [b]
df	3	4	4	4
Asymp. Sig.	.000	.001	.000	.000

Table 4.2 shows the actual result of the chi-square goodness-of-fit test. The test statistics is statistically significant for all the four groups.

(i) For JSS2: $\chi^2(3) = 22.3$, $p < .0005$. Therefore, we can reject the null hypothesis and conclude that there are statistically significant differences in the teachers' grading of the students' English proficiency, with most teachers selecting A2 ($N = 16$) compared to the other grades.

(ii) For JSS3: $\chi^2(4) = 18.9$, $p < .005$. Most teachers selected B1 ($N = 13$) compared to the other grades.

(iii) It is the same case for SS1: $\chi^2(4) = 29.8$, $p < .0005$, with most teachers selecting B2 ($N = 15$) compared to the other grades.

(iv) It is not different for SS2 class: $\chi^2(3) = 23.1$, $p < .0005$, with most teachers selecting C1($N = 14$) compared to the other grades.

In the light of these results, it was concluded that the proficiency level assigned by the majority of the raters to each group best describe their language ability. While the teachers' judgement may be subjective, this is the most reliable and practicable available option for assigning proficiency levels to these groups of learners. As a result of this, the four sub-corpora in this study will henceforth be referred to as NILECORP-A2 (66, 545 words), NILECORP-B1 (73,246 words), NILECORP-B2 (128,613 words) and NILECORP-C1 (248,513 words) while they will collectively still be referred to as Nigerian Learner Corpus of English.

4.4.3 The Strengths and Weaknesses of CEFR

The strength of CEFR levels lies in its transparency and coherence in that the descriptors are flexible and inclusive. This means it can be applied across different languages more readily. Even among teachers of the same language in similar contexts, there can be a lot of variety in the descriptions of language proficiency levels. Obviously, this variability increases significantly across difference languages in different context. The CEFR makes it easier to view language proficiency levels reliably and with shared understanding.

However, the CEFR is not unproblematic. As North (2014: 23) puts it, CEFR scales represent a heuristic of 'scaled teacher perceptions' and as such it does not necessarily reflect second language development. The framework's descriptors were calibrated in the so-called Swiss Project (Council of Europe, 2001; Wisniewski, 2017). First, it consisted of roughly 2,000 descriptors from a range of test of English; and then 300 language teachers, in a series of workshops, sorted the descriptors according to the category of L2 communicative competence they perceive the descriptors belong (Wisniewski, 2017). Another major criticism of this entirely teacher-based scaling perspective is that the descriptors were not matched onto empirical learner language to see if the teachers' perceptions correspond to authentic learner behaviour (Wisniewski, 2017). Yet another constraint is that learners in the framework's levels are placed according to skills-based rather than knowledge-based criteria (Council of Europe, 2013). Being skills-based means the levels are imprecise (Milton & Alexiou, 2009). The implication of this is that it is possible to place learners at several of the CEFR levels. Milton (2010: 229) argues that "users of the system often find it difficult to match learners or materials to the levels with any precision and different people, different examiners, even different national examination systems, can apply the CEFR's levels descriptors very differently". This, as Milton and Alexiou (2009: 194) puts it, "potentially devalue the framework and diminishes its usefulness". If it had been knowledge-based, it probably would have been clearer on the vocabulary size of each level.

Milton (2010) argues that the linking of linguistic features of performance to the CEFR levels looks to be important, and the use of vocabulary size measurements, and the tests to derive such measurements would help users of the system in different schools or countries apply

grading criteria more consistently and confidently. Although the Common European Framework is not explicit on vocabulary, Milton's (2010) study shows progressively higher vocabulary scores are associated with progressively higher levels in the CEFR hierarchy. However, there is individual variation and overlap between the scores that learners attain within the CEFR Levels. This, once again, highlights the problem of assigning proficiency levels to a whole group. But despite the criticism of the CEFR, it remains the best framework available in the sense that it evaluates students' language competence from broader dimensions. It provides a comprehensive description of the language skills of each proficiency level. The reason for using the CEFR framework in this study is to ensure the four proficiency levels involved in the study are clearly defined in a way that will be understandable to researchers and whoever is reading the thesis who may not be familiar with the Nigerian context. By assigning a clearly defined proficiency levels to the corpus texts, an important design criterion in computer learner corpora compilation has not just been met, but this study will now have relevance that transcends the Nigerian context as researchers elsewhere may now use it for learner corpus research comparing learner groups. With the study corpus now in place, I will now discuss the reference corpora used in this research.

4.5 The Reference Corpora

A reference corpus is a corpus of text which is used as a standard for comparative purposes. In selecting a reference corpus, Goh (2011: 239) notes that "genre and diachrony are more important factors to consider than other factors […] especially in that the differences in these two factors, unlike those in other factors such as corpus size and varietal difference, bring about significant difference in the number of the keywords". Keywords are those words whose "frequency is exceptionally high (positive keywords) or low (negative keywords) in comparison with a reference corpus" (Xiao & McEnery, 2005: 68). Using corpus text of similar genre as a reference corpus means both corpora (the reference corpus and the study corpus) will significantly have similar textual patterns.

Four corpora were considered as a possible reference corpus in this study. One of them is the British Academic Written English Corpus (BAWE). The BAWE is a 6,506,995 words corpus of proficient university-level student writing. It contains 2,761 pieces of proficient assessed student writing, ranging in length from about 500 words to about 5000 words. It was collected as part of the project, 'An Investigation of Genres of Assessed Writing in British Higher Education' (Heuboeck, Holmes & Nesi, 2007; Alsop & Nesi, 2009). But BAWE is not suitable for this study, essentially because the texts of the corpus were written by speakers of various L1s described as 'proficient university-level students' – a description which is rather vague.

Another corpus which was considered is the British National Corpus (BNC). The 100 million words corpus is a collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written. Ninety percent of the corpus is made up of written texts which include extracts from newspapers, specialist periodicals and journals for all ages and interests, academic books, published and unpublished letters, school and university essays, among many other kinds of text. The spoken component consists of orthographic transcriptions of unscripted informal conversations and spoken language collected in different contexts (Burnard, 2007; Leech & Rayson, 2014). Although the BNC is a native English corpus, the extremely diverse genres covered in the corpus and its enormous size makes it unsuitable for this purpose.

The third corpus which was considered is the Louvain Corpus of Native English Essays (LOCNESS). LOCNESS is a 324,304 words corpus of native English essays made up of British pupils' A level essays (60, 209 words), British university students' essays (95, 695 words) and American university students' essays (168, 400 words). There are 430 essays in the corpus, 317 of them which represents 228, 501 words are argumentative essays while the other 113 essays are expository essays, literary texts and text on literature but mostly rather argumentative. Some of the essays were timed and the writers had no access to reference tools. Others were either not rigidly timed or not timed at all and reference tools were used (CECL Louvain, 2015). LOCNESS and NILECORP are similar in many ways. They are both compiled in a similar context – academic context; they are both written by young

students; and the texts are similar genres which mean they may have similar textual patterns. In view of these factors, LOCNESS was chosen as the primary reference corpora.

The fourth corpus that was considered is the Corpus of Global Web-Based English (GloWbE). GloWbE is a corpus of World Englishes which contains about 1.9 billion words of text from twenty countries. The twenty countries include: United States, United Kingdom, Canada, Ireland, Australia, New Zealand, India, Sri Lanka, Pakistan and Bangladesh. The other countries are: Singapore, Malaysia, Philippines, Hong Kong, South Africa, Nigeria, Ghana, Kenya, Tanzania and Jamaica. The Nigerian component of GloWbE contains 42.6 million words drawn from 37, 285 web pages from 5, 520 websites and blogs (Davies, 2013).

There is no information on the English language proficiency of the writers of the texts. But it is plausible to conclude that they will be English speakers of varying proficiency, probably ranging from intermediate to advanced speakers of English. The Nigerian component of GloWbE is the largest corpus of Nigerian English. This mega sub-corpus is, therefore, chosen as the secondary reference corpus for this study. Any instantiation of word combination that is not found in the collocation dictionaries and the native reference corpus will be looked up in the Nigerian component of GloWbE before labelling them as deviant collocations. But if such word combinations are found in the corpus, they will be regarded as Nigerian English collocations. With all the corpora (study corpus, primary reference corpus and secondary corpus) in place, I will now proceed to the extraction of collocational candidates from the study corpus and the primary reference corpus. In the next section, I will explain how LOCNESS was used in this study.

4.6 Procedures and Analytical Approach

The section details the approach and procedures of extracting collocational candidates form the corpora. It also discusses the analytical approach adopted for the data analysis.

4.6.1 Extraction of Collocational Candidates

This study initially intended to investigate six lexical collocations but had to reduce it to two lexical collocations after seeing the sheer amount of work it would require to investigate that many. The two sub-types of collocations this study investigates are the Verb Noun and Adjective Noun collocations. There is no known publication that has investigated these collocations at this scale, particularly from the perspective of World Englishes. I started with the analysis of the LOCNESS as collocations in this corpus are used as the baseline for comparisons with the collocations in the learner corpus. Due to the wide range of collocations involved in this study, there are six main steps in the procedure of the native corpus analysis.

1. Using the word list function of AntConc, I scanned the corpus for all the nouns in it and created a frequency list for them. Any noun that appears five times and above were isolated. The cut-off point of five is arbitrary but it is necessary for ease of analysis. And this does not in any way imply that such nouns are necessarily frequently or less frequently used in our day-to-day language use.
2. I used the same corpus analysis software to create concordances for each of the nouns so that all the instances of Verb + Noun combinations could be extracted.
3. Subsequently, all instances of co-occurrence of the above combination within a collocational span of up to five words to left hand side of the nouns, being the key word, were regarded as collocational candidates and were, therefore, extracted accordingly. All the extracted combinations were checked in the Oxford Collocations Dictionary for Students of English (McIntosh, 2009) and The BBI Dictionary of English Word Combinations (BBI). These two dictionaries were used because the former is a corpus-based dictionary while the latter was used because collocations in this study were based on BBI classification. If the combinations were listed as collocation in either of the dictionaries, they were noted as collocations. Various L2 collocational studies have used similar procedure of verification (Nesselhauf, 2005; Wang & Shaw, 2008; Laufer & Waldman, 2011).
4. Following the same approach in procedure 1 above, I scanned the corpus for all adjectives in it and created a frequency list for them with the cut off set on five instances as above.

5. I created a concordance for each of these adjectives so that all the instances of Adjective + Noun combinations within the collocational span of five words to the right-hand side of the adjective being the keyword could be identified and extracted.
6. All the extracted collocational candidates were checked in the dictionaries, and if any instances of Adjective + Noun combination that was listed as a collocation in either of the dictionaries was isolated.

Upon completion of the collocational candidates, every combination which was not listed in the dictionaries was not used in this study. They were regarded as open/free combination and are not part of the object of this research.

4.6.2 Extraction of Collocational Candidates for the Nigerian Learner Corpus

The next step is the analysis of the learner corpus and then the analysis of its four sub-corpora. The analysis proceeded in a way that was similar to that of the native speaker corpus. First, all the nouns and adjectives found in the native speaker corpus were extracted from the learner corpus. These structures were identical in both corpora, but beyond this, I also extracted all the other collocational combinations in the learner corpus which are not in the native speaker corpus. Most of the existing comparative studies on L2 collocation competence and development did not account for the other collocations (whether correct or erroneous collocations) which are in the learner corpus but not in the reference corpus. This study, however, included all other nouns and adjectives which were in the learner corpus but not in the native speaker corpus because they obviously indicated something about the learners' collocational competence. So, using the same wordlist function of AntConc, I scanned the learner corpus for all the nouns and adjectives and isolated them.

The next step is also similar to the native speaker corpus analysis. I created concordances for the nouns and every instance of Verb + Noun combinations were identified. Then all well-formed combinations were verified in the two dictionaries. The same procedure was used to extract the adjectives from the learner corpus. Subsequently, all the Adjective + Noun combinations were verified in the dictionaries. All the combinations that were not found in

the dictionaries were noted for further analysis. Because this study investigates collocations from the perspective of World Englishes, all the collocational candidates that were not found in the dictionaries were checked up in the Nigerian component of GloWbE. If found in the Nigerian component of GloWbE, they were included in the study and regarded as Nigerian English collocations. The four sub-corpora (NILECORP-A2, NILECORP-B1, NILECORP-B2 and NILECORP-C1) that made up Nigerian Learner Corpus of English (NILECORP) were analysed separately. This was to enable me compare NILECORP-C1 (the most proficient of the four groups) with LOCNESS (the native speaker corpus) and to do comparisons between the four sub-corpora representing the proficiency levels.

4.6.3 Data Analytical Approach

This thesis answers four broad questions with several sub-questions under each of them. This section describes briefly how the data are analysed to answer the research questions. The first question investigates the extent to which native and non-native writers make use of collocations. To answer this question, the data from LOCNESS are compared with the data from NILECORP-C1 which is the most proficient of the four learner groups. It starts with a comparative analysis of the numbers of verb noun and adjective collocations in both corpora, and then compares the collocations produced by the two groups in terms of the linguistic complexity of their constituents. It also compares the number of figurative collocations (collocations with additional nuances and associations) produced by both groups and concludes by investigating the number of congruent and incongruent collocations produced by the NILECORP-C1 group to determine whether they use more congruent than incongruent collocations. The second main research question investigates the relationship between frequency of and exposure to input in L2 learners' speech community and their production of collocations. The correlation between the collocations produced by the NILECORP-C1 group and the frequency data on these collocations from the Nigerian component of the 1.9 billion words Global Web-Based English Corpus is investigated to determine if the learners produce more of the most frequently used collocations in Nigeria their speech community.

The third main question investigates the relationship between proficiency and the production of collocations. This question is answered through a comparative analysis of all the collocations produced across the four proficiency levels. It investigates the overall number of collocations produced by each group and determines the percentage of congruent and incongruent collocations produced at each proficiency level. It also investigates the relationship between proficiency and the use of linguistically complex verb noun collocations in terms of the collocation span and the structural properties of their constituents by comparing the verb noun collocations extracted from the four sub-corpora. Finally, on this question, the relationship between proficiency and the use collocations with additional nuances and associations by comparing the degree of semantic opacity and transparency of the collocations produced by the learners. The last main question is concerned with the nature and causes of the errors in the collocations produced by the learners. It starts with the identification, classification and the analysis of all the erroneous verb noun and adjective noun collocations extracted from the four sub-corpora. Some element of contrastive interlingual analysis is used to identify the types of collocations which are the most problematic for the learners, the nature and causes of the collocational errors in the four sub-corpora. It also investigates the similarities and differences in the error across the four proficiency levels. Finally, the collocational errors are analysed to determine what proportion of the collocational errors are due to: [a] Inter-lingual factors and [b] Intra-lingual factors. My knowledge of Yoruba language – my L1 which is also the L1 of the participants in this study – will be brought to bear in the error analysis. What counts as collocational error in this study is not premised on the notions of norms and standards of some of the prestigious varieties of English but on the basis of the acceptability in Nigerian English language usage context. There is a further discussion on this later in the thesis.

Chapter Five

Native Speakers and L2 Learners' Use of Collocations

5. 0 Introduction

This chapter investigates and compares the extent to which native and non-native writers make use of collocations in a written text and considers a number of research questions: (1) Do native speakers use more collocations than L2 learners? (2) What is the difference between the collocations produced by the two groups in terms of the linguistic complexity of their constituents? (3) How many semantically burdensome collocations are produced by both groups? And (4) How many congruent and incongruent collocations are produced by the L2 learners? It provides a detailed comparative analysis of all the Verb Noun and Adjective Noun collocations produced in the LOCNESS and NILECORP-C1 (this learner group is equivalent to the Common European Framework of Reference for Languages proficiency level C1). I initially wanted to investigate five sub-sets of collocations (Verb + Noun, Adjective + Noun, Adverb + Adjective, Verb + Adverb, Noun + Verb, and Noun + Noun) but because of the volume, I decided to reduce it to two sub-sets – something manageable which I will be able to investigate in-depth.

It is divided into five sections as follows:

The first section presents the overall descriptive statistics of the data used for this study. It includes numbers of tokens in the two corpora, the numbers of verb noun and adjective noun collocations extracted from the corpora, the semantically burdensome collocations produced by both groups, and the statistics on the congruent and incongruent collocations produced by the learners.

In the second section, I will go beyond statistical data to qualitative analysis by identifying, comparing and interpreting evidence from the various collocational expressions produced in the corpora. This section focuses on the comparative analysis of the linguistic complexity of the verb noun collocations produced by the native speakers and the L2 learners. By linguistic

complexity, I mean the complexity in terms of the collocation span and the structural properties of the constituents of the verb noun collocations. This section is divided into two sub sections. One sub-section focuses on the collocation span while the other focuses on the structural properties of the constituents of the collocations.

In the third section, the data on 'semantically burdensome' collocations will be presented and analysed. As I have said earlier, the collocations in the data set I refer to as 'semantically burdensome' are essentially metaphorical collocations. Metaphorical collocations are "imbued with a bewildering range of connotative and associative meanings" (Phillip, 2011: 26). They could be problematic for L2 learners because of a double meaning. In such instances, meanings have been modified to introduce additional nuances and associations (Phillip, 2011). The aim of this analysis is to understand the extent to which L2 learners produce and use semantically opaque collocations with varying degree of idiomaticity. The surface wording of these types of collocations does not reflect the meaning of the whole (Ibid). Using the term 'semantically burdensome' collocations seems to account for the continuum of opacity in this type of collocations rather than using the term 'semantically opaque collocations' which does not seem to account for this continuum. The 'weight' of the production/processing burden is dependent on the degree of opacity within the continuum.

In the fourth section, I will analyse the data on the congruent and incongruent collocations produced by the learners. The congruent collocations have the same conceptual bases and linguistic expressions in both English and Yoruba while incongruent collocations are collocational expressions that are totally different conceptually and linguistically in the two languages. This section is divided into two sub-sections. The first sub-section focuses on congruent and incongruent verb noun collocations while the second sub-section focuses on congruent and incongruent adjective noun collocations.

Finally, at the end of this chapter, there will be a discussion section where I will interpret and explain my findings and examine whether and how my research questions have been answered. In this section, I will explain any new understanding or insights about the problems that have been investigated after taking the findings into consideration. The

discussion will show how my findings relate to the immediate literature on native speakers and L2 learners' use of collocations. It will also explore the theoretical significance of my findings as well as outline any new areas for future research which my findings have suggested.

5. 1 Overall Results

The descriptive statistics presented in this section describe the basic features of the data used to investigate the extent to which Native Speakers and L2 Learners make use of collocations in their written texts. It provides simple summaries of the samples and measures used in this section. Two corpora were used at this stage – LOCNESS and NILECORP-C1. LOCNESS has 326,838 word tokens and 16, 185 word types while NILECORP-C1, the most proficient group of the four learner groups used in this thesis, has 252,003 word tokens and 9,193 word types. All the nouns involved in the study appear at least six times in the corpora and only verb + noun and adjective + noun collocations that occur twice and above were included in the analysis. All nouns that appear fewer than six times and all instances of verb + noun and adjective + noun collocations that appear fewer than two times were excluded from the analysis. Based on these criteria, 711 verb noun and 740 adjective noun collocations were extracted from LOCNESS while 1,847 verb noun and 531 adjective noun collocations were extracted from NILECORP-C1.

The first step in this statistical data analysis is to check whether the data are appropriate for the comparative analysis. In order to manage the data properly, the presence of outliers must be detected, investigated and addressed. Outliers are unusual points in the data that differ substantially from the other observations (Barnett & Lewis, 1994). These outliers, if undetected and addressed, could potentially skew the results leading to mistaken conclusions and inaccurate predictions.

In the verb noun collocations extracted from the learner corpus, three structures have a frequency that differs substantially from the other observations. These structures are: 'keep

bad company' which appears 839 times, 'go/went + school' which appears 125 times, and 'keep + friend' which appears 90 times in the NILECORP-C1. These figures are extremely far apart in comparison to other structures in the corpus which appear between 2 and 46 times. It is important to investigate the reason for the outliers. An investigation reveals that while giving out the essay writing task to the learners in the course of compiling the corpus, the accompanying note on how to write the essay contains the expressions: 'keeping bad company' and 'keeping bad friends'. This must have influenced the learners' usage of the expressions. It is concluded that the setting of the essay title must have influenced their use of the structure 'go/went + school' as they were writing about keeping bad friends in school. In view of the above, these three collocational structures which were produced 1,054 times were therefore excluded from this analysis. All these outliers were excluded from the data. Only the verb + noun collocations in the NILECORP-C1 were affected.

With the outliers out of the way, the overall count of the verb noun collocations in the NILECORP-C1 is 793 well-formed verb noun collocations. Considering the size of the corpus, that translates to 0.31% (793 ÷ 252,003 × 100). In comparison, overall count of verb noun collocations in the Native speaker corpus is 711 representing 0.21% (711 ÷ 326,838 × 100) considering the size of the corpus. An independent sample t-test was conducted to compare the native speakers and the L2 learners' verb noun collocations. There was no significant difference in the number of verb noun collocations used in LOCNESS (M = 7.48, SD = 9.78) and NILECORP-C1 (M = 8.94, SD = 10.06); t (183) = 0.997, p = 0.320 [Cohen's d: 0.14]. A total of 528 adjective noun collocations were extracted from the learner corpus which accounts for 0.20% (528 ÷ 252,003 × 100) while 740 adjective noun collocations represent 0.22% (740 ÷ 326,838 × 100) were extracted from the native corpus. An independent t-test comparing the two groups reveals there is no significant difference in the scores for LOCNESS (M = 6.98, SD = 10.35) and NILECORP-C1 (M = 8.80, SD = 10. 27); t (164) = 1.090, p = 0.277.

A total of 1,324 collocations being the combination of the verb noun and adjective noun collocations were extracted from NILECORP-C1 represent 0.52% considering the size of the learner corpus. In comparison, 1,451 – which is the combination of verb noun and adjective noun collocations – extracted from LOCNSS represent 0.44%. In proportion to the size of the corpora, the learners produced more collocations than the native speakers. An

independent t-test comparing the two groups also reveals there is no significant difference in the scores for LOCNESS (M = 7.22, SD = 10.06) and NILECORP-C1 (M = 8.88, SD = 10. 11); t (349) = 1.52, p = 0.128. Looking at the data purely in numerical terms, the learners produced more collocations than the native speakers. See table 5.1 below for more details.

Table 5.1 Statistics of the data used in this chapter

	LOCNESS	**NILECORP C1**
Word Tokens	326,838	252,033
Word Types	16,185	9,193
V + N Collocations	711	793
Adj + N Collocations	740	531
Semantically Burdensome (V + N)	103	88
Semantically Burdensome (Adj + N)	38	26
Congruent V +N Collocations	N/A	448
Incongruent V + N Collocations	N/A	345
Congruent Adj +N Collocations	N/A	387
Incongruent Adj + N Collocations	N/A	144

But if we consider the number of verb noun collocational structures produced against the frequency, the native speakers produced 96 verb noun collocational structures collectively

used 711 times. By verb noun collocational structure, I mean a string of verb and the co-occuring (collocating) noun. For example, the structures: 'evoke a degree of sympathy', 'evoke any sympathy' and 'evoke more sympathy' will count as one 'evoke + sympathy' verb noun collocation structure used three times in the corpus. But the structures: 'have any sympathy' and 'feel some sympathy' will count as two separate verb noun collocation structures ('have + sympathy' and 'feel + sympathy') used once each in the corpus. The focus of the analysis at this stage is to investigate the selection of the right co-occuring element and not the internal structures. So, if we divide the number of collocational structures by the overall frequency of usage multiplied by hundred (96 ÷ 711 × 100), that would be 13.5% while the learners produced 89 verb noun collocational structures collectively used 793 times which is 11.2%. The 96 verb noun collocational structures produced by the native speakers, if divided by 326,838 (the size of the corpus) multiplied by hundred is 0.02% while the 89 structures produced by the learners is 0.03%. Considering the numbers of verb noun collocational structures produced by both groups in proportion to the size of the respective corpus, the learners produced slightly more verb noun collocations than the native speakers. However, the learners repeated several of the collocational structures in their text than the native speakers. They seem to overuse five structures: 'express + view' was used 46 times, 'have + friend' was used 45 times, 'give + birth' was used 42 times, 'lead to + trouble' was used 40 times and 'take + care' was used 39 times. The native speakers also appeared to overuse the structure 'have + children' – it was used 83 times in the corpus. The native speakers' over usage of this structure may be because there is hardly any other way of expressing the concept of 'having children' without using the verb 'have'. In the case of the learners, while there seems to be a limited alternative way of expressing the concepts of 'having friend' and 'taking care' without using the verb 'have' and 'take' respectively, this is not the case with 'expressing view', 'give birth' and 'lead to trouble'. They are a clear case of over use. These five structures alone were used 212 times in the learner corpus contributing so much to the overall frequency data of the learners' usage of collocations in their text.

As for the adjective noun collocations, the native speakers produced 107 adjective noun collocation structures collectively used 740 times in the corpus while the learners produced 60 structures collectively used 531 times. If we divide the number of adjective noun collocational structures by the overall frequency of usage multiplied by hundred (107 ÷ 740 ×

100), that would be 14.4%. The learners' 60 adjective noun collocational structures if divided by the overall frequency of usage multiplied by hundred (60 ÷ 531 × 100) is 11.2%. The 107 adjective noun collocational structures produced by the native speakers, if divided by 326,838 (the size of the corpus) multiplied by hundred is 0.03% while the 60 structures produced by the learners using the same calculation is 0.02%. Considering the numbers of adjective noun collocational structures produced by both groups in proportion to the size of the respective corpus, the native speakers produced slightly more adjective noun collocations than the L2 learners.

To have a clearer picture of how many collocations the two groups produced in their texts, there is a need to look at the combination of the structures (verb noun and adjective noun collocational structures) in proportion to the size of the respective corpus. The native speakers produced 203 adjective noun and verb noun collocational structures which if divided by 326,838 (the size of the corpus) multiplied by hundred translates to 0.06% while the learners produced 149 adjective noun and verb noun structures which, using the same calculation, translates to 0.05%. What this means is, based purely on frequency of usage regardless of how many times a particular structure is repeated, the L2 learners produced slightly more collocations (0.52% against 0.44%) in their text than the native speakers. But if we consider the numbers of different collocational structures produced, the native speakers produced slightly more collocations (0.06% against 0.05%) than the L2 learners.

So, based on the parameter set for the extraction of collocations from the corpora and considering the size of the corpora, the learners produced almost equal numbers of collocations in their text to what the native speakers produced. In answering the first research question above, native speakers do not necessarily produce more collocations in their text than L2 learners. This finding is seemingly counter-intuitive. It raises a few questions which will be addressed later when analysing the linguistic complexity of the collocations produced by the two groups. In quantitative terms, the first notable finding here is that relatively advanced learners (CEFR – C1 equivalent) of English from an English as a second language context where the learners have frequent exposure to the input outside the classroom, in this instance, have shown that they can produce as many collocations in a written text as native speakers do.

Having said this, the descriptive analysis shows that the native speakers produced a total of 22 semantically burdensome collocations – 14 verb noun and eight adjective noun semantically burdensome collocations. The L2 learners on the other hand produced a total of seven semantically burdensome collocations – six verb noun and one adjective noun semantically burdensome collocations. Clearly, the data reveals L2 learners use fewer semantically burdensome collocations in their written text. This will be analysed further in section three.

A further analysis of the 793 verb noun collocations produced by the learners reveal that 448 representing 56.4% of them are congruent while 345 representing 43.6% are incongruent. Out of the 531 adjective nouns collocations which were extracted from the learner corpus, 387 representing 72.8% are congruent while the other 144 representing 27.2% are incongruent – they have no equivalent in the Yoruba language. An in-depth analysis of this finding and what it means in terms of the collocational proficiency of the learners will be presented in section four.

5. 2 Linguistic Complexity of Verb Noun Collocations

In the section above, it was established that (in quantitative terms) relatively advanced learners of English from an English as a second language context where the learners have frequent exposure to the input outside the classroom could produce as many collocations in a written text as native speakers do. From this section, I will now go beyond statistical data to qualitative analysis by identifying, comparing and interpreting evidence from the various collocational expressions produced in the corpora. This section answers the research question: What is the difference between the collocations produced by the two groups in terms of their linguistic complexity. By linguistic complexity, I mean the complexity in terms of the collocation span, and the structural properties of the constituents of the verb noun collocations. But only the verb noun collocations will be analysed for the linguistic complexity. The analysis of the linguistic complexity of the verb noun collocations begins with the analysis of their collocation span. The span is called ‘collocation window’ (Brezina,

McEnery & Wattam, 2015: 140); and the collocation window for this study is set for L5 – R5. It refers to the distance between the node and the collocate. By node, I mean the element being studied while the element that co-occurs in the defined environment of this node is the collocate. In this analysis, all bigram collocations are excluded. Bigrams are two-word collocations – just the node and the collocate without any lexical element in between. Also excluded from the analysis are all three-word collocations that have demonstrative adjective, definite and indefinite article, and possessive determiner between the collocate and the node.

5.2.1 Collocation Span

The data reveals that the native speakers group and the learners group each produced 46 three-word collocations. However, while the native speakers produced 120 four and five-word collocations, the learners only produced 59 such structures. In total, the native speakers produced 163 long span collocations (three to five-word collocation) collectively used 197 times. The learners on the other hand produced a total of 102 long span collocations collective used 191 times. An independent t-test comparing the two groups reveals there is significant difference in the scores for LOCNESS (M = 1.20, SD = 0.75) and NILECORP-C1 (M = 1.84, SD = 1. 75); t (164) = -4.101, p = 0.001. This result confirms the learners' productive knowledge of long span collocations significantly lags behind that of native speakers. Any collocational structure that has, at least, one different lexical element between the collocate and the node were included in the study. For example, 'make an important decision', 'make a hard decision', 'make a firm decision' and 'made the right decision' count as four collocations. But in the analysis in section 5.1 above, all these collocations were calculated as one 'make + decision' collocation which was used four times. They are counted as four different collocations here because the focus of the analysis is on the internal elements. See below some examples of the long span collocations produced by both groups:

LOCNESS	**NILECORP-C1**
come to such biased conclusions	take my own decision
come to the conclusion	make good decision
	making bad decision
draw the wrong conclusion	make a lot of difficult decisions

draw their own conclusion
draw her own conclusion
draw totally the wrong conclusion

evoke a degree of sympathy
evoke any sympathy
evoke more sympathy
evoke both sympathy and …
have any sympathy
feel some sympathy

making enough profit
make any profit
making such a healthy profit
make more of a profit

make a strong argument
make an effective argument
makes their entire argument
make their whole argument
make a firm, decisive argument
make for an effective argument
make an ever-stronger argument
make a much more effective argument

heed my teacher's advice
heed to my friend's advice
heed to a good advice

make quick money
make some money
make a lot of money

accept their offer
accept the dangerous offer
accept all their offers

rushed to the hospital
rushed him to the hospital
rushed her to the nearby hospital

make bad friend
made many friend
make new friend
make boy friend
make two new friends

The findings in this sub-section reveal that: (1) while relatively advanced learners of English could produce as many collocations in a written text as native speakers do, they produce fewer long span collocations. (2) Considering the number of long span collocations against the frequency, learners seem to repeat certain collocations in their text more than the native speakers. (3) Considering the number of long span collocations and their frequency of usage in relation to the overall number of verb noun collocations produced by the learner and their frequency of usage in the corpus, learners tend to have preference for two-words collocations like ‘make decisions’ as opposed to ‘make a lot of difficult decisions’. (4) Native speakers overwhelmingly produce more long span collocations than L2 learners. What this means is that the nature of collocations, in terms of the span, produced by native speakers in written texts is noticeably different from the ones produced by relatively advanced L2 learners of English.

5.2.2 Structural Properties of the Verb Noun Collocations

Having established that native speakers produce more long span collocations than L2 learners, the next phase in the analysis of the linguistic complexity of the verb noun collocations produced by the two groups is the structural properties of the collocations. The focus of this analysis is on the collocations that have collocations in their constituents (collocations within collocations). All the long span collocations produced by both groups were analysed and all the verb noun collocations that have collocations within their structures were isolated. The data reveal that out of the 163 long span verb noun collocational structures the native speakers produced 44 structures which are collectively used 55 times have collocations within their structures. It is clear from the number of structures versus the frequency of usage that the native speakers did not over use any of these collocations. See table 5.1 for the details of all the collocations that have collocations within them.

Table 5.1 Collocations within Verb Noun Collocations in LOCNESS

Collocation within Verb Noun Collocations in LOCNESS			
draw the *wrong conclusion*	2	make an *important decision*	2
draw totally the *wrong conclusion*	1	make a *hard decision*	1
		make a *firm decision*	1
take *full advantage*	2	made the *right decision*	2
have *easy access*	2	achieve his *ultimate goal*	2
making such a *healthy profit*	1	made a *clear statement*	1
		making such *bold statement*	1
have *disastrous consequences*	1		
have *dire consequences*	1	take such *drastic action*	1
have *harmful consequences*	1	take the most *appropriate action*	1
have *serious consequences*	1		
have *disastrous* global *consequences*	1	earning *sums of money*	1

		earning a *lot of money*	2
bear the *additional burden*	1	making a *big sum of money*	1
		make a *lot of money*	3
tell her *life story*	1	make the *amount of money*	1
tell the *whole story*	1	make that *big money*	1
create a better *transport system*	1	accept *total responsibility*	1
create a new *road system*	1	take *moral responsibility*	1
introduce one *monetary system*	1	bear *moral responsibility*	2
introduce a heavy *tax system*	1	shoulder the *moral responsibility*	1
introduce some type of *tolling system*	1		
		make a *strong argument*	1
commit a *serious crime*	2	make an *ever-stronger argument*	1
convicted of *drug crimes*	1	*weaken considerably* the arguments	1
watch *daytime television*	1	making *spelling errors*	1
		Total	**55**

There are 13 structures in this table, collectively used 55 times in LOCNESS

In comparison, out of the 102 long span verb noun collocations produced by the learners, only 13 structures collectively used 32 in the corpus have collocation within them. An independent t-test comparing the two groups reveals there is significant difference in the scores for LOCNESS (M = 1, SD = 0.001) and NILECORP-C1 (M = 0.3947, SD = 0. 49536); t (80) = 8.113, p = 0.001. This shows there is a wide gap between the structural properties of the collocations produced by native speakers and relatively advanced L2 learners of English. See table 5. 2 below for the details of the collocation within collocations produced by the learners:

Table 5.2 Collocations within Verb Noun Collocations in NILECORP-C1

Collocation within Verb Noun Collocations in NILECORP-C1					
take *good care*	7	keeping *late night*	1	pay *less attention*	1
make *quick money*	2	keeping *good record*	2	spend a *lot of money*	1
make a *lot of money*	1				
		write a *short story*	9	learn a *good lesson*	1
heed to a *good advice*	1	write this *short story*	4		
narrated the *whole story*	1	give *medical treatment*	1	Total	32

There are 13 structures collectively used 37 times in NILECORP-C1

Considering the total number of collocations with collocations within them produced by the learners versus the total number of times these structures are used in the corpus (13/32), the data shows the learners used few collocations several times. This could be the case of overusing certain favourite collocations. While this group of learners were able to produce an almost equal number of verb noun collocations as the native speakers in their written text, there is a wide gap in the structural complexity of their collocations. The main findings here are: (1) there is a huge gap in terms of the structural complexity of the constituents of verb noun collocations produced by native speakers and relatively advanced learners of English from an English as a second language context where the learners have frequent exposure to the input outside the classroom; (2) second language learners tend to overuse a few favourite structurally complex verb noun collocations.

5.3 Analysis of Semantically Burdensome Collocations

This section, which is divided into four sub-sections, identifies and analyses the extent to which the learners and the native speakers produce semantically burdensome collocations – collocations on the upper end of the continuum of semantic opacity. The focus in this section is on the semantic properties of collocations. This is an aspect that has been, hitherto, neglected in collocation research. The collocations in this category might entail more cognitive load to process by the L2 learners because, to a varying degree, their meanings have been modified to introduce additional nuances and associations (Phillip, 2011). Some of them have delexical verbs which establish their meaning from the words (in this case, the nouns) they are combined with. According to McCarthy (2014), collocations that are formed around these verbs are unpredictable and hard to recall when needed and as a result difficult even for advanced learners to produce. Others are more metaphorical in which at least one constituent of the collocation is applied to an object or action to which it is not literally applicable.

There are a total of 250 instances of semantically burdensome collocations out of the 2,775 collocations extracted from the two corpora. The learners produced 109 semantically burdensome collocations – 83 verb noun and 26 adjective noun collocations while the native speakers produced 141 semantically burdensome collocations – 103 verb noun and 38 adjective noun collocations. This means 8.2% of all the collocations produced by the L2 learners are semantically burdensome while 9.7% of the collocations produced by the native speakers are semantically burdensome. An independent t-test comparing the two groups reveals there is no significant difference in the scores for the L2 learners (M = 6.41, SD = 9.09) and the native speakers (M = 5.54, SD = 4. 54); t (46) = 0.95, p = 0.34. The only difference is the degree of opacity of the semantically burdensome collocations produced by both groups. If put in a single continuum within the same processing system from fully transparent to fully opaque, most of the ones produced by the native speakers would be on the upper end of opacity while most of the one produced by the learners would be on the lower end of opacity. This may be attributable to the cognitive load of processing semantically burdensome collocations. The cognitive load varies depending on the degree of opacity of the structure. These findings on the production of semantically burdensome collocations mirror findings in the study of other similar linguistic phenomenon like the productive and receptive

knowledge of L2 metaphors and idioms (Doiz & Elizari, 2013; Zibin & Hamdan, 2014; Zibin, 2016). These findings bring to bear the effect of semantic features in the acquisition of L2 collocations. This will be discussed further in the discussion section at the end of this chapter.

In order to find out where the difference really lies in the semantically burdensome collocations produced by both groups, I will do a fine-grained analysis of this phenomenon from four angles focusing on their semantic properties and the degree of opacity of the structures produced. To this end, the first sub-section will focus on the semantically burdensome verb noun collocations produced by the learners, the second section will focus on adjective collocations they produced, the third sub-section will focus on the verb noun collocations produced by the natives and the fourth sub-section will examine the adjective noun collocations produced by the natives.

5.3.1 Analysis of Semantically Burdensome V + N Collocations in NILECORP-C1

The learners produced 10 different collocational structures which have varying degree of semantic opacity. These 10 structures were collectively used 83 times in the learner corpus. See table 5.3 below for more details:

Table 5.3 Semantically Burdensome Verb Noun Collocations in NILECORP-C1

Semantically Burdensome V + N Collocations in NILECORP-C1			
Collocations	Frequency	Collocations	Frequency
take + care	39	tarnish + reputation	2
take + bath	10	turn down + offer	2
tarnish + image	9	shed + blood	2
fall in + love	7	keeping + late night	2
handle + equipment	7		
damage + image	3	**Total**	**83**

Some of the structures are clearly metaphorical and idiomatic, and have added new shades of meaning. For instance, the structures: 'tarnish + image', 'tarnish + reputation', 'damage + image' have additional nuances and associations. The verb 'tarnish' in the company of reputation or image here does not mean losing lustre but referring to the denting of one's reputation. It might not be easy for L2 learners to grasp the meaning of this type of expression. Similarly, the surface wording of the collocations 'fall in + love' and 'turn down + offer' do not reflect the meaning of the whole.

The delexical verb 'take' in 'take + care' and 'take + bath' takes on new meaning when used with bath and care. *Bath* and *care* are not things you 'take' but they are things you 'do' but it sounds awkward and unacceptable to say, '*I want to 'do' good care of you instead of I want to take good care of you or to say I want do my bath instead of I want to take my bath'*. Also, saying 'shed + blood' instead of 'kill' and 'handle +equipment' instead of 'use + equipment' makes them problematic for learners ('shed + blood' could be ambiguous out of context but in Nigerian English, it often means killing someone and the context will clearly suggests that). However, if I were to put these collocations in a single continuum within the same processing system from fully transparent to fully opaque, I would not consider them to be fully opaque. Notwithstanding, it would require a considerable cognitive effort for learners to produce these types of collocations. The key finding here is that the learners did not use fully opaque verb noun collocations.

5.3.2 Analysis of Semantically Burdensome Adj + N Collocations in NILECORP-C1

The learners produced seven different collocational structures which have varying degree of semantic opacity. These seven structures were collectively used 26 times in the learner corpus. See table 5.4 below for more details:

Table 5.4 Semantically Burdensome Adjective Noun Collocations in NILECORP-C1

Semantically Burdensome Adj + N Collocations in NILECORP-C1			
Collocations	Frequency	Collocations	Frequency
bright + future	13	sweet + experience	2
innocent + blood	3	deep + voice	2
strong + influence	2	tight + security	2
bright + student	2	**Total**	**26**

They produced very few semantically burdensome adjective noun collocations with varying degree of opacity. The collocation 'innocent + blood' is fully opaque. In Nigerian English, if we say, for instance, 'the government must put an end to the shedding of innocent blood' The 'innocent blood' in that statement refers to someone/people without guilt of a crime or offence. It is hard to guess the meaning from the surface words. The 'bright' in 'bright + student' and 'bright + future' has nothing to do with reflection of light but being clever and a promising future. In the same manner, 'deep' in 'deep + voice' has nothing to do with depth (like two or three feet deep) but sound. While these collocations may be problematic for L2 learners, they are not fully opaque apart from 'innocent + blood'. We can see here again that L2 learners seem to avoid fully opaque collocations.

5.3.3 Analysis of Semantically Burdensome V + N Collocations in LOCNESS

This sub-section focuses on the analysis of the semantically burdensome verb noun collocations produced by the control group. There are 103 instances of semantically burdensome verb noun collocations in the native speaker corpus. I will analyse some of them to determine how different they are, in semantic terms, from the ones extracted from the learner corpus.

Table 5.5 Semantically Burdensome Verb Noun Collocations in LOCNESS

Semantically Burdensome V + N Collocations in LOCNESS			
Collocations	Frequency	Collocations	Frequency
take + responsibility	15	shoulder + responsibility	3
take + advantage	13	face + risk	2
take into + account	11	take + revenge	2
bear + burden	7	bring up + child	2
take + action	7	introduce a **heavy tax** system	1
tackle + question	6	committing **intellectual suicide**	1
take + risk	6	commits **symbolic suicide**	1
face + problem	5	committing **philosophical suicide**	1
take + decision	5	make a **strong argument**	1
take + notice	4	making such **bold statement**	1
take + precaution	4	making such a **healthy profit**	1
bear + responsibility	4	**Total**	**103**

In comparison with the ones produced by the learners, if I put the semantically burdensome verb noun collocations produced by the native speakers in a single continuum within the same processing system from fully transparent to fully opaque, I would put some of them toward the upper end of fully opaque. For instance, collocations like: 'introduce a heavy tax system', 'make a strong argument', 'making such bold statement' and 'making such a healthy profit' have elements within them that have a high degree of opacity. We can see how the native speakers have used tax as though it is something that has weight, but the meaning has nothing to do with physical weight. They combined 'argument' with 'strong', 'statement' with 'bold' and 'profit' with 'healthy' to convey metaphorical meaning. In these instances, meanings have been modified to introduce additional nuances and associations (Phillip, 2011). A greater cognitive process is involved in producing such expressions. Though the learners produced a substantial number of collocations, in comparison with the native speakers, they have produced very few collocations that have these characteristics.

Another example of collocations with additional shades of meaning can be seen in the way the native speakers used the word suicide. While it may not be semantically burdensome for L2 learners to produce 'commit + suicide', it could be semantically challenging for them to produce: 'committing philosophical suicide', 'commits symbolic suicide' and 'committing intellectual suicide' as the natives have done. Other similar examples are their usage of 'tackle + question' as if dealing with the question in a physical combat; 'bear + burden', 'bear + responsibility' and 'shoulder + responsibility' – all of which have metaphorical meanings. Additionally, they used more collocations with delexical verbs as can been seen on the table above. One key finding at this point is that while this group of Yoruba-speaking learners of English have produced almost as many verb noun collocations as the natives did, however, there is a big gap in the semantic quality of the collocations produced by both groups. By which I mean the usage of collocations to reflect various shades of meaning from fully transparent to fully opaque.

5.3.4 Analysis of Semantically Burdensome Adj. + N Collocations in LOCNESS

There are 38 instances of collocations that belong to this category in the native speaker corpus. They have varying degree of opacity. Some of them might be problematic for learners to produce. 'naked + truth' for instance, has nothing to do with being naked rather it means plain unadorned facts, without concealment or embellishment. We can see how far removed is the meaning from the words. The word 'strong' as the collocates of position, argument, evidence and opinion is not a reference to having power. We can see here that the word strong as used with the nodes have implied meaning. Consider 'powerful + emotion' for instance, while it might not be very difficult for L2 learners to understand the meaning [depending on their level of proficiency], it could be cognitively challenging for learner to produce this type of collocation. See the table 5.6 below for all the semantically burdensome adjective noun collocations extracted from the native speaker corpus.

Table 5.6 Semantically Burdensome Adjective Noun Collocations in LOCNESS

Semantically Burdensome Adj. + N Collocations in LOCNESS			
Collocations	Frequency	Collocations	Frequency
strong + argument	20	strong + position	2
weak + argument	4	strong + evidence	2
naked + truth	3	powerful + emotion	2
strong + opinion	3		
intellectual + suicide	2	**Total**	**38**

There are eight semantically burdensome adjective noun collocational structures in this table, they were collectively used 38 times.

5.3.5 Summary of Findings on Semantically Burdensome Collocations

Overall, the analysis of these collocations based on their semantic opacity yields some important findings:

- 8.2% of all the collocations produced by the L2 learners are semantically burdensome.
- 9.7% of the collocations produced by the native speakers are semantically burdensome.
- If the semantically burdensome collocations produced by the learners were to be put in a single continuum within the same processing system from fully transparent to fully opaque, they would be on the lower end of opacity.
- If the semantically burdensome collocations produced by the native speakers were to be put in a single continuum within the same processing system from fully transparent to fully opaque, they would be on the upper end of opacity.

In summary, there is a gap between the collocations produced by the learners and the native speakers in terms of using collocations to reflect various shades of meaning from fully transparent to fully opaque.

5.4 Analysis of Congruent and Incongruent Collocations

This section answers the research question: how many of congruent and incongruent collocations are produced by L2 learners? It focuses on the analysis of the collocations produced in the Learner Corpus based on cross-linguistic relationships and differences. The collocations that have lexical components that are similar in Yoruba and English are regarded as congruent while the ones with lexical components that are different in the two languages are incongruent (Yamashita and Jiang, 2010). The effect of L1 on the acquisition of L2 collocation has been the subject of various studies (Biskup, 1992; Siyanova & Schmitt, 2008; Yamashita & Jiang, 2010; Wolter & Gyllstad, 2011; Laufer & Waldman, 2011; Phoocharoensil, 2012) with evidence suggesting that learning incongruent collocations is problematic. In view of this, this comparative analysis of the extent to which natives and non-natives use collocations in their written text goes further to investigate the extent to which the learners use both congruent and incongruent collocations. This section, as has been stated earlier, is divided into two sub-sections. The first sub-section focuses on congruent and incongruent verb noun collocations while the second sub-section focuses on congruent and incongruent adjective noun collocations. I will now present the data on congruent and incongruent verb noun collocations.

5.4.1 Congruent and Incongruent Verb Noun Collocations

A total of 89 verb + noun collocational structures were extracted from the Learner Corpus C1. These collocational structures were collectively used 793 times by the learners. Out of the 89 verb + noun collocational structures, 40 of them are incongruent representing 44.9% while the other 49 structures representing 55.1% are congruent verb noun collocations. The 40 incongruent verb + noun collocational structures were used 345 times in the learner corpus. The congruent structures on the other hand were used 448 times. An independent sample t-test was conducted to compare the frequency of usage of congruent and incongruent verb noun collocations. There is no significant difference in scores for incongruent verb noun collocations (M = 8.57, SD = 11.10) and congruent verb noun collocations (M = 9.18, SD = 9.25); $t(87) = -0.27$, $p = 0.77$.

5.4.1.1 Incongruent Verb Noun Collocations

In this section, I will do a fine-grained analysis of some of the incongruent verb noun collocations produced by the learners. The table below shows all the incongruent verb noun collocations.

Table 5.7 Incongruent Verb Noun Collocations

Incongruent Verb Noun Collocations			
	Frequency		Frequency
express + view	46	fetch + water	4
give + birth	42	acquire + knowledge	4
take + care	39	sentence + prison	3
make + friend	30	impart + knowledge	3
make + money	20	form + habit	3
keep + gang	10	contract +disease	3
take + bath	10	achieve + dream	3
write + examination	9	say + prayer	3
tarnish + image	9	make + love	3
bear + children	9	lavish + money	2
perform + task	9	keeping + late + night	2
commit + crime	8	draw + attention	2
take + decision	8	catch + attention	2
handle + equipment	7	turn down + offer	2
fall in + love	7	keep + record	2
damage + image	6	obtain + money	2
heed + advice	6	break + regulation	2
pay + attention	6	tarnish + reputation	2
achieve + goal	5	engage + prostitution	2
pamper + children	5		
narrate + story	5	Total	345

There are 40 verb noun collocation structures in this table, and they were used 345 times.

All these collocational expressions are totally different in the two languages; they do not have equivalent in the Yoruba language. They are so different to the extent that if some of them are considered in isolation from their context, a Yoruba learner of English (depending on their English proficiency level) may misunderstand their meaning. For instance, the expressions 'fall in love', 'handle equipment', 'tarnish image' and 'take bath' are far removed from the way we would express these concepts in Yoruba language. If a Yoruba learner of English were to express these concepts with cross-linguistic influence, for 'fall in love' they would probably produce something like 'full of love', for 'handle equipment' they will say 'use equipment'. To say 'use equipment' is correct and congruent with the Yoruba equivalent expression. It should have been much easier for the learners to say this but instead they opted for 'handle equipment' which is incongruent. The structure 'fall in love' is figurative; which further makes it far removed from its Yoruba equivalent. Despite this, the learners correctly produced this collocation seven times in the learner corpus.

For incongruent collocations like 'tarnish image' Yoruba-speaking learners of English due to L1 interference, will probably say *'destroy your reputation'* and for 'take bath' they may say *'do bath'.* However, despite being incongruent, the learners produced these collocations acceptably without any negative transfer. For the other collocations in this data, the learners could have produced *'do + birth'* for give birth, *'do + care'* for take care, *'do + decision'* for make/take decision, *'say + view'* for express view, *'choose + friend'* for make friend (using 'choose' in the context where 'make' is the appropriate verb), *'do + crime'* for commit crime and *'do + examination'* for write examination. Again, the learners demonstrated their ability to produce incongruent collocations. Some of them were produced quite frequently in the corpus. For instance, 'express view' was produced forty-six times, 'give birth' was produced forty-two times, 'take care' was produced thirty-nine times, and make friend was produced thirty times.

Such collocations as 'break regulation', 'keep record', 'turn down offer', 'draw attention' 'keep late night', 'form habit' and 'make love' should normally be problematic for Yoruba learners of English because the combinations do not have Yoruba equivalent. A typical learner with Yoruba as L1 will most likely say *'disobey regulation', 'write record', 'reject an*

offer', 'call attention', 'walking late night', 'learn habit', and *'do love'.* These would be the direct translation of the English collocational structures into the Yoruba way of expressing these concepts. But the data shows these learners produced incongruent verb noun collocations 343 times correctly. One would have expected they would produce far fewer incongruent verb noun collocations. But that is not the case here which is remarkable because the two languages belong to two different linguistic families: Yoruba is Niger-Congo language (Campbell, 1991) while English is Germanic a member of the wider Indo-European language. It is important to find out why it seems these learners do not have much difficulty producing these many incongruent verb noun collocations despite the apparent lack of similarity in the two languages. The key could lie in the context in which they learn English. And the level of exposure they might have had to these incongruent verb noun collocations in their speech community might have enhanced their ability to produce incongruent collocations. In the next chapter, I will attempt to investigate why the learners have successfully produced so many incongruent verb noun collocations. Meanwhile, some of the issues raised above will be dealt with extensively in chapter eight which is dedicated to collocational error analysis. But for now, I will proceed to analyse the learners' production of congruent verb noun collocations.

5.4.1.2 Congruent Verb Noun Collocations

Collocational expressions that have the same conceptual bases and linguistic expressions in both L1 and L2 (congruent) are thought to be less problematic for L2 learners (Bahns, 1993, Nesselhauf, 2005). This group of learners seem to have confirmed that. As expected the learners used more of collocational expressions that are congruent with the Yoruba language. They produced 48 different verb + noun collocational structures which were used 448 times in the learner corpus. See table 5.8 below for a list of all the verb noun collocation structures.

Table 5.8 Congruent Verb Noun collocations in NILECORP-C1

Congruent Verb Noun Collocations			
	Frequency		Frequency
have + friend	45	fail + examination	6
lead to + trouble	40	learn + lesson	5
have + sex	27	ask + forgiveness	5
take + hospital	26	have + tendency	5
listen + advice	21	have + opportunity	5
write + story	19	keep + secret	5
choose + friend	18	look + trouble	5
cause + problem	16	pick + phone	5
spend + money	15	Resume + school	5
tell + story	13	pay + money	4
rush + hospital	12	collect + result	4
have + accident	11	burst + tears	4
have + dream	11	follow + rule	3
have + experience	11	accept + offer	3
pass + examination	9	give + instruction	3
have + problem	8	Collect + result	3
go + hospital	8	shed + tears	3
give + treatment	8	Sentence + prison	3
drive + car	7	hold + position	2
cause + trouble	7	shed + blood	2
make + trouble	7	Fight + war	2
bring + shame	7	follow + instruction	2
make + decision	7	send + prison	2
read + story	7	give + punishment	2
		Total	448

Forty eight different verb noun collocations structures were used 448 times in the corpus.

It appears to be quite easy for the learners to produce congruent collocations. 448 out of the 793 verb noun collocation produced which is 56% have Yoruba equivalent. Even such expression as 'shed blood' which is idiomatic was not difficult for the learners to produce. This is obviously because of the congruence of the concept and the linguistic expression in both Yoruba and English. To shed blood is not just to cause blood to flow but in most cases means to kill some by violence except the context clearly suggests otherwise. Considering the degree of idiomaticity of the expression, one would expect this might be problematic for them. But this expression has direct equivalent with similar degree of idiomaticity in Yoruba. 'Shed blood' in Yoruba, literally is *'ta eje s'ile'* (*'ta'* is shed, *'eje'* is blood and *'si ile'* is on the ground).

shed	*blood*	*[on the ground]*
ta	*eje*	*s'ile*

And this, in Yoruba, means to kill someone violently. The surface meaning of the Yoruba expression, just like the English, is to spill blood on the ground. While the Yoruba language adds 'on the ground' to that collocation, English does not but implicitly, when blood is shed it would be on something. While this expression seems to be conceptually congruent, the form does not perfectly map on to its English language equivalent. The words *'shed'* and *'blood'* [*'ta'* and *'eje'*] do map on perfectly to their Yoruba translation equivalent but the Yoruba equivalent of the collocation explicitly includes *'s'ile'* [on the ground]. The question here is, can this type of collocation be categorized as congruent? There seems to be a gap in the literature on the theoretical concept of collocational congruency. There seems to be no explicit criteria for dichotomous congruency classification (Lee & Lin, 2013). The notion of congruency, which could be subjective, has mostly depended on individual researcher's lexical knowledge and word meaning interpretation to give a binary classification of congruent and incongruent collocations. Having said that, because the Yoruba equivalent of the collocation 'shed blood' largely maps on to its English equivalent and the other part [*s'ile*] which does not seem to map on seem to exist implicitly in the English equivalent, this collocation is more of a partial congruency. The Kroll and Stewart's (1994) model which I discussed earlier does not seem to envisage cases of partial conceptual congruency. It is, however, categorized as congruent because collocations are traditionally categorised as either congruent or incongruent and this is clearly not incongruent. I will discuss this further in the

discussion chapter when discussing the overall findings of this study within Kroll and Stewart' Revised Hierarchical Model of bilingual language processing.

Meanwhile, another example of L1 mediating in the production of L2 collocation is the expression 'hold + position'. The learners used the collocation 'holding an important position' twice in the corpus. The verb 'hold' both in English and Yoruba means, in most case, to gasp something with your hand. And as such, a position – an abstract concept – is not something that can be gasped with one's hand. But despite this ambiguity, the learners were able to produce and use it appropriately. The verb 'dimu' (hold) in Yoruba also mean to be in a position (as in office). In Yoruba syntax, that verb can be spit to accommodate lexical elements in between. So, the Yoruba equivalent of 'hold important position' is *'di ipo pataki mu'* [*ipo* is position, *pataki* is important, *dimu* is to hold]. If literally translated, it would be *'hold position important'.*

hold	*important*	*position*
dimu	*pataki*	*Ipo*

Though the Yoruba syntax for this collocation is not congruent with its English equivalent, the concept and the linguistic elements are congruent. Another example of a collocation that is not syntactically congruent produced by the learner is: 'choose godly friends'. But the collocate 'choose' and the node 'friends' are congruent. The verb choose is *'yan'* in Yoruba and the node friends is *'ore'* while the lexical element 'godly' which comes between the collocate and the node is *'to n'iwa bi Olorun'* in Yoruba [literally means 'having the quality of God]. Syntactically, the word 'godly' will come after the node friends and not before it as the case in English. So, the direct equivalent of that collocational structure in Yoruba is 'choose friends godly' [*yan ore to n'iwa bi Olorun*].

choose	*godly*	*Friend[s]*
yan	*to n'iwa bi Olorun*	*Ore*

Note that while the collocate 'choose' has a one-word equivalent in Yoruba and the node 'friends' also has a one-word equivalent in Yoruba, the lexical element 'godly' which is a constituent of this collocation does not have a one-word equivalent in Yoruba. However,

there is an equivalent concept in Yoruba. Despite this, the learners were able to select the correct collocate and node as well as re-arrange the syntax correctly.

The key findings so far are (1) This group of relatively advanced learners produced high numbers of incongruent verb noun collocations (2) These learners produced more congruent verb noun collocations than the incongruent ones. (3) The learners do not seem to have difficulty producing and using appropriately collocations that are idiomatic if they congruent (4) The learners can produce lexically congruent collocations that are not syntactically congruent. There was no evidence that syntactic incongruence was an issue in the production of such collocations. I will now consider adjective noun collocations.

5.4.2 Congruent and Incongruent Adjective Noun Collocations

This section which is divided into two sub-sections focuses on the production of congruent and incongruent adjective noun collocations. The congruent and incongruent adjective noun collocations are identified and analysed. Some of these collocations are peculiar to Nigerian English and are hardly used in Native English (British English). A total of 60 adjective + noun collocational structures were extracted from the learner corpus. These structures were used 531 times in the corpus. Twenty two out of the 60 adjective + noun structures are incongruent while the other 38 structures are congruent. These 22 incongruent structures were used 144 times in the learner corpus while the 38 congruent structures were used 387 times. A cursory look at this frequency data reveals that this group of learners produced more congruent adjective noun collocations than incongruent ones. An independent t-test comparing the number of times congruent and incongruent adjective collocations were used in the corpus reveals there is no significant difference in scores for incongruent adjective noun collocations (M = 6.40, SD = 6.23) and congruent adjective noun collocations (M = 10.20, SD = 11.71); t (59) = -1.40, p = 0.16. In the following sub-section, I will analyse the incongruent and congruent adjective noun collocations produced by the learners in more details.

5.4.2.1 Incongruent Adjective Noun Collocations

The learners managed to produce 144 incongruent adjective noun collocations representing 26.1% of all the adjective noun collocations produced in the learner corpus. This may mean the learners have difficulty producing incongruent adjective noun collocations and therefore, opted for congruent collocations. See table 5.9 below for all the incongruent adjective noun collocational structures and the number of times each of them was used in the corpus.

Table 5.9 Incongruent Adjective Noun Collocations in NILECORP-C1

Incongruent Adjective Noun Collocations			
Collocations	Frequency	Collocations	Frequency
secret + cult	28	last + night	4
bright + future	13	average + student	3
elder + brother	12	curious + mind	3
long + time	11	strong + influence	2
bad + disposition	11	pipe-borne + water	2
bouncing + baby	11	varying +degree	2
unwanted + pregnancy	10	tight + security	2
moral + education	6	deep + voice	2
nearest + future	5	bright + student	2
nonchalant + attitude	5	quick + reaction	2
social + miscreant	4		
peer + pressure	4	Total	144

As the frequency data suggests, the first seven structures seemed overused in the corpus. The notion of overuse and underuse as characteristics of learner language are matters of frequency. If learners use a word or structure more frequently than native speakers, such word or structure may be regarded as overused. If they use a word or structure more frequently instead of other words or structures which may be used in the same context, that too may be regarded as overuse (Kamshilova, 2017). These structures are regarded as overused on the basis of the latter. The structures 'secret + cult' and 'bright + future' were used 28 and 13 times respectively. And the structures 'elder + brother' was used 12 times while the structures 'long + time', bad + disposition' and 'bouncing + baby' were each used 11 times. The seventh structure 'unwanted + pregnancy' was used 10 times. These seven structures account for 96 out of the 141 times that incongruent adjective noun collocations were used in the corpus. That is 68% of the incongruent adjective noun collocations. This seems to confirm previous findings that L2 learners overuse a narrow range of collocations (Durrant & Schmitt, 2009).

Among the incongruent adjective noun collocations produced by the learners, there are some combinations which are particular to Nigerian English. The combination of 'social' and 'miscreant', for instance is hardly used in native British English (and other prestigious varieties of English). Social miscreants are people like drug addicts and pushers, alcoholics, thieves; people who often foment trouble in the streets and other unsocial behaviours. This collocation was used four times in the learner corpus. There is no Yoruba equivalent word for 'social + miscreant'. The concept of social miscreant is a relatively new phenomenon in Nigerian big cities, yet this group of Yoruba-speaking English learners were able to produce this extremely incongruent collocation. A search for this collocation on the 1.9 billion words Corpus of Global Web-Based English (GloWbE) reveals this expression is mainly used in Nigeria and Ghana. The learners were probably able to produce it because of exposure to the collocation in their speech community.

Another collocation in this list is 'nonchalant + attitude'. This collocation was used five times in the corpus. A search for this collocation on the GloWbE comparing how frequently this expression is used across the 20 countries corpus reveals it is almost exclusively used in Nigerian English. But more interestingly, the learners used another collocation with a

different meaning. The combination 'average + student' in Nigerian English, among other meanings, means a student who is neither the best nor the worst in terms of academic performance. See the extract from the learner corpus below:

*... an expensive school. She happened to be an **average student** and all effort made to improve her...*
*... continuous assessment test. Although I was an **average student** but a drop in point in any of...*
*... our time. Since he knew Reuben was an above **average student** and that come rain or shine he...*

All the three instances of the collocation in the corpus are used in reference to performance. What all this mean is that learners' productive knowledge of collocation is influenced by the variety of English they are exposed to. I will expand on this in the discussion chapter. However, there is one combination among the incongruent adjective noun collocations which the learners frequently used in their text but does not appear in the Nigerian component of GloWbE. The learners produced 'bad + disposition' 11 times though it was incongruent and apparently fewer frequently used in Nigeria. To sum up, (1) these learners produced fewer incongruent adjective noun collocations. (2) They overused a narrow range of incongruent adjective noun collocations. (3) The learners' choice and meaning of collocations is influenced by the variety of English they are exposed to.

5.4.2.2 Congruent Adjective Noun Collocations

As expected the learners produced more adjective noun collocations which can easily be matched with their Yoruba conceptual and linguistic equivalent than the incongruent ones. But even then, there is evidence of overuse. Out of the 38 congruent adjective noun structures that were produced, 10 of them appeared to be overused. The following collocations have unusually high frequency in relation to the other collocations on the list: 'good + friend' (56 times), 'bad + behaviour' (40 times), 'bad + character' (30 times), 'peer + group' (28 times), 'best + friend' (27 times) and 'fellow + student' (23 times). Others that seem to be overused are: 'bad + attitude (17 times), 'bad + influence' (15 times), 'armed + robbery' (14 times) and 'bad + habit' (10 times). The syntax of Yoruba language is very different from English. The above adjective noun collocations have the adjective come before the nouns in English but in Yoruba, the adjectives come after the nouns. These structures are not syntactically congruent. But the volume of the congruent adjective noun collocations produced by the learners regardless of the fact that they are not syntactically

congruent suggests that syntactic incongruence does not make it problematic for them to produce. See the list of the congruent adjective noun collocation below for more details:

Table 5.10 Congruent Adjective Noun Collocations

Congruent Adjective Noun Collocations			
Collocations	Frequency	Collocations	Frequency
good + friend	56	second + wife	5
bad + behaviour	40	important + position	5
bad + character	30	bad + manner	5
peer + group	28	bad + reputation	5
best + friend	27	good + record	5
fellow + student	23	major + problem	5
bad + attitude	17	little + brother	4
bad + influence	15	good + character	4
armed + robbery	14	hard + work	4
bad + habit	10	true + friend	3
good + manner	9	several + times	3
final + examination	8	innocent + blood	3
best + student	7	age + group	3
class + mate	7	sweet + experience	2
good + care	6	secret + society	2
bad + language	6	evil + companion	2
close + friend	5	tight + friend	2
good + news	5	bad + news	2
main + purpose	5		
big + shame	5	**Total**	**387**

There are 38 structures used 387 times in the corpus.

5.4.2.3 Summary of Findings on Congruent and Incongruent Collocations

Overall, this analysis shows there are 1, 324 instances of adjective noun and verb noun collocations (793 verb noun collocations and 531 adjective noun collocations). Of this number, 835 are congruent (448 congruent verb noun collocations and 387 congruent adjective noun collocations) representing 63.1% of all the collocations produced. 489 representing 36.9% of all the collocations produced are incongruent (345 incongruent verb noun collocations and 144 incongruent adjective noun collocations). An independent t-test shows there is no significant difference in the number of times incongruent collocations are produced (M = 7.88, SD = 9.62) and the number of times congruent collocations produced (M = 9.70, SD = 10.55); $t(146) = -1.08$, $p = 0.28$.

In summary, the key findings on this section are:

- 36.9% of all the (adjective noun and verb noun) collocations produced by the learners are incongruent while 63.1% are congruent – these L2 learners seem more inclined to using congruent collocations than incongruent collocations.
- This group of relatively advanced learners produced considerably high numbers of incongruent verb noun collocations – 44% of the verb noun collocations produced.
- These learners produced more congruent verb noun collocations than the incongruent ones – 56% of the verb noun collocations produced.
- The learners do not seem to have difficulty producing and using appropriately verb noun collocations that are idiomatic if they congruent.
- The learners can produce congruent verb noun collocations that are not syntactically congruent.
- These learners produced fewer incongruent adjective noun collocations - 26.1% of the adjective noun collocations produced.
- The learners produced more congruent adjective noun collocations than the incongruent ones – 73.9% of the adjective noun collocations produced.
- They overused a narrow range of incongruent adjective noun collocations.
- The learners' choice and meaning of collocations is influenced by the variety of English they are exposed to.

5.5 Discussion

Four main themes were investigated in this chapter. The first research questions investigated the quantity of collocations in L2 learners' written text versus native speakers' written text while the second investigated the linguistic complexity in terms of collocation span and structural complexity of the constituents of the verb noun collocations produced by L2 learners versus native speakers. The third research question inquired into L2 learners versus native speakers' ability to use collocations to convey various shades of meaning ranging from fully transparent to fully opaque, and the fourth question investigated the learners' production of congruent and incongruent collocations.

The first finding in this chapter is apparently counter-intuitive. The literature on L2 collocational competence and development (Granger, 1998; Nesselhauf, 2005; Siyanova & Schmitt, 2008; Barfield & Gyllstad, 2009; Laufer & Waldman, 2011; Henriksen, 2013) indicates L2 collocations deficiency is a pervasive phenomenon in second language acquisition and as such one would expect L2 learners to use fewer collocations in their written text in comparison to native speakers. On the contrary, in quantitative terms, the first notable finding is that relatively advanced learners (CEFR – C1 equivalent) of English from an English as a second language context where the learners have frequent exposure to the input outside the classroom, in this instance, have shown that they can produce as many collocations in a written text as native speakers do. Considering the numbers of verb noun and adjective noun collocations extracted from the two corpora in proportion to the size of each corpus, the native speakers did not significantly produce more collocations than the L2 learners. Based purely on the frequency of the instances of collocations regardless of how many times a particular structure is repeated, the L2 learners produced slightly more collocations (0.52% against 0.44% in relation to the size of each corpus) in their text more than the native speakers. But if we consider the numbers of different collocational structures produced, the native speakers produced slightly more collocations (0.06% against 0.05%) than the L2 learners. While L2 collocation is actually problematic for learners, the difference in the collocations produced by relatively advanced learners of English and native speakers does not necessarily lie in the quantity of collocations produced but in the linguistic complexity of the collocations. This is what seems to be missing in the literature. A large body of research already existed on the knowledge and use of collocations by L2 English

learners and this has been further expanded by the readily availability of learner corpus. With computer corpora firmly established as a research tool, the field Learner Corpus Research has, among other things, broadened our knowledge of collocations and the difficulties learners have producing them. However, we seem to have focused too much on learners' ability to select the appropriate co-occurring words.

As far back as 1998, Howarth (1998: 36) claims that "the problem facing the non-native writer or speaker is knowing which of a range of collocational options are restricted and which are free". According to him "the ability to manipulate such clusters [collocations which are partly restricted] is a sign of true native speaker competence and is a useful indicator of degrees of proficiency across the boundary between non-native and native competence (ibid: 38). He argues that "learners' difficulties lay chiefly in differentiating between combinations that are free and those that are somehow limited in substitutability" (ibid: 42). In the last three decades or so, this has been the focus of many studies. The literature seems to be saturated with studies investigating learners' ability to select appropriate co-occurring words. But beyond selecting the appropriate co-occurring words, which other difficulties do learners have with collocational competence and development? This leads us to the next theme that was investigated in this chapter.

One aspect that appears to have been neglected in the literature is the linguistic complexity of the collocations produced by L2 learners. Linguistic complexity in terms of the span of the collocations produced by learners in comparison to the ones produced by native speakers, and the structural complexity of the constituents of verb noun collocations produced by L2 learners. This study has revealed that native speakers overwhelmingly produce more long span collocations than L2 learners. Most of the verb noun collocations produced by the learners are bigrams [two words collocations]. While native speakers also produced many bigrams, they however, distinctively produced far more long span collocations than the learners. What this means is that the nature of collocations, in terms of the span, produced by native speakers in written texts is remarkably different from the ones produced by relatively advanced L2 learners of English. What does this mean in terms of L2 collocational competence and development? Looking beyond the node and collocate and learners' ability to select appropriate co-occurring words in collocational research could give us a better

insight into the nature of collocations produced by L2 learners. As this study reveals, while this group of learners were able to produce almost an equal numbers of verb noun collocations as the native speakers in their written text, there is, however, a wide gap in terms of the structural complexity of the constituents of the verb noun collocations produced by native speakers and this relatively advanced L2 learners of English. The native speakers produced many collocations that have collocations within them. This reflects the extent of formulaic language in native speaker texts. But the collocations produced by the learners did not have as many rich lexical elements. What this means in terms of SLA and development of fluency is that L2 learners' inability to sufficiently produce long span collocations with formulaic expressions within them may stand in the way of fluency. Various studies have "established that formulaic language provides processing advantages and is essential for using language fluently and idiomatically, both for native and non-native speakers" (Gonzalez & Schmitt, 2015: 1). But how much do L2 learners use collocations to convey idiomatic meaning in their written text? This leads us to the third theme which inquired into L2 learners versus native speakers' ability to use collocations to convey various shades of meaning ranging from fully transparent to fully opaque.

For too long, L2 collocational research has neglected the learners' ability to use collocation to convey various shades of meaning from fully transparent to fully opaque. Collocations have often been perceived as being semantically transparent in comparison to other formulaic expression like idioms. But this cannot be taken to mean that their meaning is always a compositional function of the meanings of their constituents (Trantescu, 2015). It is difficult to establish which of the constituents contributes which proportion of the meaning of the collocation. This makes the semantic aspects of collocations hard to capture except by studying them within their wider textual and domain context (Bartsch, 2004). Perhaps, this explains why this aspect has not received much attention. When Bartsch (2004: 72 - 75) was characterising collocations in terms of their semantic transparency, she identified four possibilities as follows:

(1) All constituents of the collocation contribute an aspect of their transparent meaning; the collocation remains semantically fully transparent in the sense that its meaning is constituted of overt realisations of one of the potential senses of each of its constituents.

(2) At least one of the constituents of the collocation does not contribute lexical meaning. One constituent may be delexicalized – losing part or all of its independent meaning.
(3) The collocation remains superficially transparent but carries an additional element of meaning that is not overtly expressed by any of its constituents.
(4) Partly opaque collocations in which (at least) one of the constituents acquires a collocation-specific meaning which it does not have outside this particular word combination.

Essentially, **s**emantic transparency in the context of collocations can be viewed as a continuum. It is the end point of a continuum of degrees of opacity (Cruse, 1986). One end of the continuum reflects a more superficial, literal correspondence and the opposite end reflects a deeper, more elusive and figurative correspondence. With this characterisation of collocation, to what extent do L2 learners' productive knowledge of collocation reflects these lexico-semantic properties of collocations? In this study, I take the position that the elements of semantic opacity of collocation would require additional cognitive burden to process and produce, hence the justification for my reference to these collocations as semantically burdensome.

As Gyllstad & Wolter (2016) rightly point out, one type of word combination for which there is a comparative lack of research in terms of processing and representation is collocation. To date, L2 collocational processing research has identified congruency and frequency of input as having definite effects (Bahns, 1993; Nesselhauf, 2005; Yamashita & Jiang, 2010; Kim & Kim, 2012; Wolter & Yamashita, 2015; González Fernández & Schmitt, 2015). However, to the best of my knowledge, in none of these studies were the semantic criteria of collocations like figurativeness or the degree of idiomaticity and semantic transparency considered in the item selection process. One study that investigates the effects of the semantic properties of collocations on their processing is carried out by Gyllstad & Wolter (2016). Using Howarth's Continuum Model to investigate free combination and collocations based on the phraseological tradition, they discovered there was a processing cost for collocations compared to free combination. This means semantic transparency affects processing of collocations but what does this mean for learners' production of L2 collocation in written form?

If there is a processing cost for collocations, then, adding the semantic properties of collocations as a factor might help us to understand its role in the production and comprehension of L2 collocations. And that is what a section of this study has attempted to do. This study has revealed that if the semantically burdensome collocations produced by the learners and the native speakers were to be put in a single continuum within the same processing system from fully transparent to fully opaque, the former would be on the lower end and the latter on the upper end of opacity. Putting this in concrete terms, learners seem to produce fewer of semantically opaque collocations. Even when they produce collocations whose semantic properties are opaque, the degree of opacity or idiomaticity is relatively low compared to what native speakers produce. If we consider this in relation to Gyllstad & Wolter's (2016) discovery that there was slower processing for collocations than free combinations, it seems that the degree of the opacity of the semantic properties of the collocations slow down the processing time. The same factor seems to have resulted in the learners in this study producing not just fewer semantically burdensome collocations but also producing collocations with less idiomaticity. Theoretically, these findings partly lend credence to the distinction made in Howarth's Continuum Model (1998). The position of the collocations in the continuum of semantic transparency/opacity is a key factor in the production of L2 collocations. Assessment of L2 collocational competence and development should, therefore, not stop at their ability to select appropriate co-occurring words but should include the ability to use collocations in various shades of meaning ranging from fully transparent to fully opaque.

Another factor which has received much attention in the literature on L2 collocational processing is congruency. Various studies have shown that congruency affects the difficulty learners have in producing and processing collocations (Bahns, 1993; Wolter & Yamashita, 2015; Peters, 2016). Many of these research findings indicate a production and processing advantage for L2 collocations that have L1 equivalent form over those that do not have equivalence even at higher levels of proficiency (Nesselhauf, 2003; Laufer & Waldman, 2011; Wolter & Gyllstad, 2011; 2013; Yamashita & Jiang, 2010). However, as Wolter & Yamashita (2015) rightly noted, it is important to point out that the idea of congruent and incongruent collocations itself is problematic to some extent because words do not always have simple and straightforward translations. In Yoruba language, for instance, the verb 'so' could be reasonably translated into 'say' or 'tell' in English. So, the concept of congruency

has its complications. Notwithstanding, there are many words in English that have Yoruba equivalent without the ambiguity described above.

In line with Yamashita & Jiang (2010) and Wolter & Yamashita (2015)'s conclusion that incongruent collocations continue to pose processing challenge to L2 learners even at higher proficiency levels, this current study also concludes that L2 learners' productive knowledge of incongruent collocations lags behind their knowledge of congruent collocations. While these learners produced almost as many collocations as the native speakers did, only 36.9% percent of the collocations they produced are incongruent. Besides, they overused a narrow range of incongruent collocations. The fact that they produced a narrow range of incongruent collocations which are then overused seems to point to the scale of the difficulty learners have producing incongruent collocations. It seems in the absence of the ability to produce incongruent collocations, the learners resorted to overusing the few ones they can produce. The key question here is how do we account for learners' deficiency in incongruent collocations in terms of L2 collocational development? Jiang's (2000) model of vocabulary acquisition could offer one way of accounting for the effect of congruency on collocational production.

Jiang's vocabulary acquisition model, which is based on an extensive review of the existing literature, proposes a three-step process for L2 vocabulary acquisition. According to this model, the first step in vocabulary acquisition consists of creating an L2 entry that is linked to a corresponding L1 word, followed by a stage where learners integrate semantic, syntactic and morphological specification into the lexical entry appropriately morphologically and phonologically/orthographically but very much remains L1-like in respect to semantics and syntax. In Jiang's view, the third stage of vocabulary acquisition is achievable through more exposure to the L2 input which will result in gradual replacement of L1-based knowledge at the lemma level with more L2-based knowledge to create a lexical entry which is "very similar to a lexical entry in L1 in terms of both representation and processing" (Jiang, 2000: 53). To account for L2 learners' production of fewer incongruent collocations, I will situate this group of Yoruba-speaking English learners somewhere in an interface between stage one and two of Jiang's lexical acquisition model. It is plausible to speculate that L2 learners start learning collocations by mapping L2 collocations into their corresponding L1collocations and "then the L2 integration stage when semantic, syntactic, morphological specifications are integrated into the lexical entry" (ibid: 47). In the absence of corresponding L1 collocations

for learners to map L2 collocation into in the case of incongruent collocations, the processing and production of L2 collocations become difficult hence their knowledge of incongruent collocations lags behind congruent collocations. I will revisit this model at the overall discussion in this thesis when I have compared the effect of congruency across different proficiency levels.

In conclusion, this chapter of the thesis has attempted to investigate the extent to which L2 learners use collocations in their written text in comparison to native speakers. The findings reveal the difference between the collocations produced by learners and native speakers does not lie in the quantity but in the linguistic complexity – structural and semantic properties of the collocations. The findings also suggest learners have difficulty producing collocations that are on the upper end of the continuum of semantic opacity and that their knowledge of incongruent collocations lags behind congruent collocations.

Chapter Six

Effects of Frequency on Collocations Production

6.0 Introduction

This chapter further elaborates the analyses reported in chapter five by considering how frequency of input affects the collocational production of L1 Yoruba learners of English. The chapter considers the following research questions: (1) What effect does the frequency of input in the Learners' speech community have on their production of verb noun collocations? (2) What effect does frequency of input in the learners' speech community have on their production of adjective noun collocations?

It is divided into six main sections as follows:

The first section provides background information on the study of the effects of frequency and exposure to input on the production of collocations. The brief overview of recent studies on the effects of frequency and exposure on L2 collocations is to set a context for this study. The second section describes the Corpus of Corpus of Global Web-Based Corpus of English (GloWbE), and the Nigerian component of GloWbE which provides the frequency data used in this investigation.

The third section which is divided into four sub-sections investigates the effects of frequency of input in the learners' speech community on the production of incongruent and congruent collocations. Using frequency data from the Nigerian component of GloWbE, the first and second sub-sections investigate the effects of the frequency of the related collocation structures in the learners' speech community on the production of frequently and less frequently used incongruent verb noun collocations produced by the learners respectively. The third and fourth sub-sections analyse the effects of frequency of input on frequently used and less frequently used congruent verb noun collocations in the learner corpus respectively.

The fourth section investigates the effects of frequency of input on incongruent and congruent adjective noun collocations using the same frequency data from GloWbE. This

section is also divided into four sub-sections. Using the same frequency data from the GloWbE, the first and second sub-sections investigate the effects of the frequency of the related collocation structures in the learners' speech community on the production of frequently and less frequently used incongruent adjective noun collocations produced by the learners respectively. The third and fourth sub-sections analyse the effects of frequency of input on frequently and less frequently used congruent adjective noun collocations in the learner corpus respectively. The fifth section presents a summary of the findings.

In the discussion, I will explain any new understanding or insights about the problems that have been investigated after taking the findings into consideration. The discussion will show how my findings relate to the immediate literature on the influence of frequency effects on the acquisition of collocations and collocation errors analysis. It will also explore the theoretical significance of my findings as well as outline any new areas for future research which my findings have suggested.

6.1 Overview of Studies on the Effects of Frequency of Collocations

This section provides a brief overview of the effects of frequency on collocations to set the context for this study. It is a widely held view in the literature that there is a close relationship between frequency and second language acquisition (Ellis, 2002a; Larsen-Freeman, 2002; Durrant & Doherty, 2010). A recent study by González Fernández and Schmitt (2015) reveals learners' knowledge of collocations correlates moderately with corpus frequency and everyday engagement with English outside the classroom. More notably, they found everyday engagement had a stronger relationship with collocation knowledge than years of English study. In another study of the effects of frequency on the processing of multiword units, the findings by Kim and Kim (2012: 838) suggest "that collocational frequency is a factor that affects the degree to which multiword units are stored as units in the mental lexicon for both native speakers and L2 learners of English. Durrant and Schmitt (2010) in a priming experiment, discovered that even one exposure to word combination resulted in a small but significant facilitation of collocation completion. In addition to these,

various other studies have concluded that frequency and exposure to input have noticeable facilitation effect (Webb, 2007; Durrant, 2008; Webb, Newton, and Chang, 2013; Peters, 2014). Some studies have suggested that "advanced learners are highly sensitive to frequency effects for L2 collocations, which seems to support the idea that usage-based models of language acquisition can be fruitfully applied to understanding the processes that underlie L2 collocational acquisition" (Wolter & Gyllstad, 2013: 451).

So, it seems high frequency and exposure to input facilitates the acquisition of collocation to some extent. But something is problematic here. How can we determine that a learner or group of learners is exposed to certain input? We cannot equate the high frequency of certain collocations in a corpus to increased exposure to those collocations. Most of the studies on the effects of frequency on collocations have used frequency data from either the Corpus of Contemporary American English (COCA) or the British National Corpus (BNC). But the frequency data from these corpora may not be representative of the learners' learning context. The assumption seems to be that if a collocation is frequent in these native speaker corpora, it may be frequent in the learners' input. Such an assumption does not take into account the learners' context and the variety of English the learners are exposed to. This current study, however, uses corpus frequency data from the learners' speech community and takes into account collocations in Nigerian English – one of the varieties of World Englishes. The collocations that are frequent in this corpus may not be frequent in COCA or BNC. So, the study investigates the effects of frequency on the learners' productive knowledge of collocations with their local context.

6. 2 Nigerian Component of Corpus of Global Web-Based English (GloWbE)

The GloWbE – a relatively new corpus released in 2013 – is composed of 1.8 billion words in 1.8 million web pages from 340,000 websites in 20 different English-speaking countries. About 60% of the corpus comes from informal blogs, and the rest from a wide range of other genres and text types (Davies & Fuchs, 2015). The large volume of the informal blogs in this corpus makes it truly representative of the variety of English used in the learners' context. The large size and the architecture of the corpus as well as its interface mean it is possible to

search and examine various linguistic phenomena across twenty varieties of Englishes. It provides frequency data for each of the twenty countries represented in the corpus.

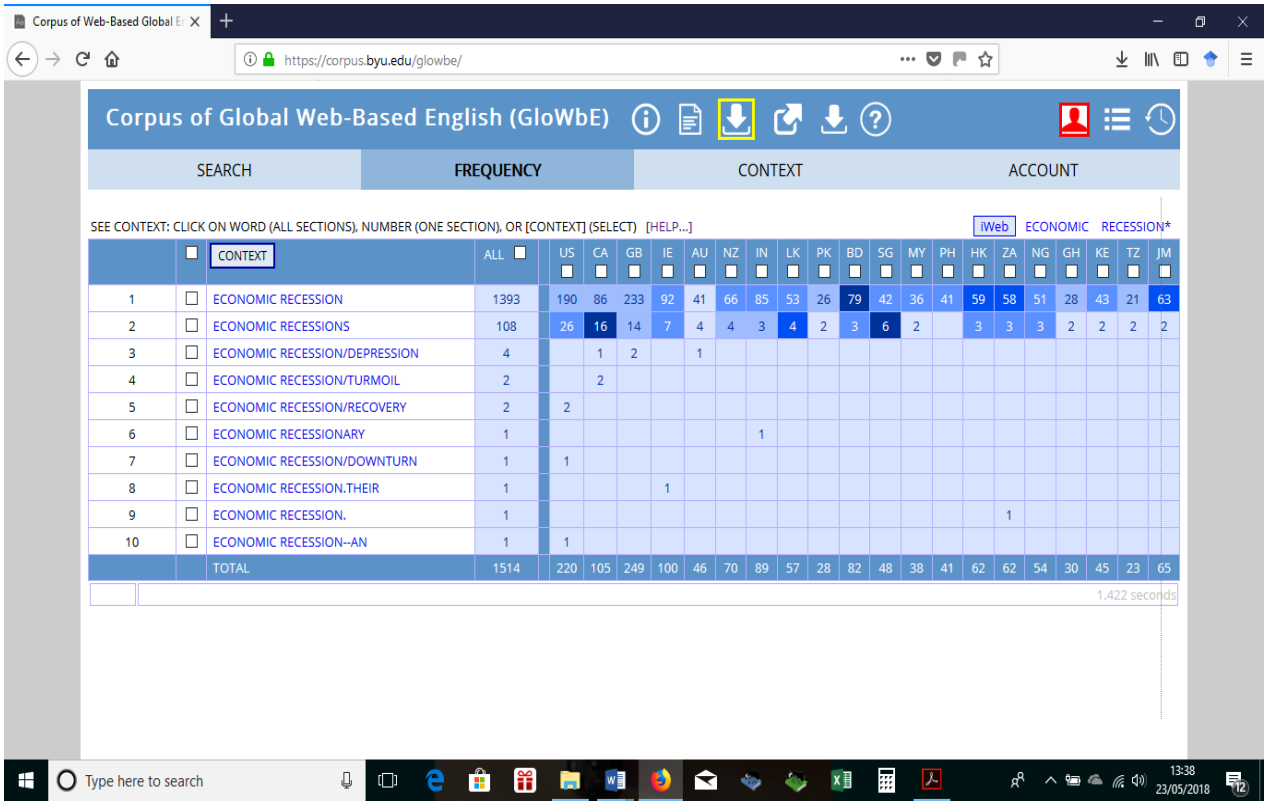
Table 6.0 Statistics of the Corpus of Global Web-Based English

Country	**Code**	**General** (may also include blogs)			(Only) **Blogs**			**Total**		
		Web sites	Web pages	**Words**	Web sites	Web pages	**Words**	Web sites	Web pages	**Words**
United States	US	43,249	168,771	**253,536,242**	48,116	106,385	**133,061,093**	82,260	275,156	**386,809,355**
Canada	CA	22,178	81,644	**90,846,732**	16,745	54,048	**43,814,827**	33,776	135,692	**134,765,381**
Great Britain	GB	39,254	232,428	**255,672,390**	35,229	149,413	**131,671,002**	64,351	381,841	**387,615,074**
Ireland	IE	12,978	75,432	**80,530,794**	5,512	26,715	**20,410,027**	15,840	102,147	**101,029,231**
Australia	AU	19,619	81,683	**104,716,366**	13,516	47,561	**43,390,501**	28,881	129,244	**148,208,169**
New Zealand	NZ	11,202	54,862	**58,698,828**	4,970	27,817	**22,625,584**	14,053	82,679	**81,390,476**
India	IN	11,217	76,609	**68,032,551**	9,289	37,156	**28,310,511**	18,618	113,765	**96,430,888**
Sri Lanka	LK	3,307	25,310	**33,793,772**	1,672	13,079	**12,760,726**	4,208	38,389	**46,583,115**
Pakistan	PK	3,070	25,852	**38,005,985**	2,899	16,917	**13,332,245**	4,955	42,769	**51,367,152**
Bangladesh	BD	4,415	30,813	**28,700,158**	2,332	14,246	**10,922,869**	5,712	45,059	**39,658,255**
Singapore	SG	5,775	28,332	**29,229,186**	4,255	17,127	**13,711,412**	8,339	45,459	**42,974,705**
Malaysia	MY	6,225	29,302	**29,026,896**	4,591	16,299	**13,357,745**	8,966	45,601	**42,420,168**
Philippines	PH	6,169	28,391	**29,758,446**	5,979	17,951	**13,457,087**	10,224	46,342	**43,250,093**
Hong Kong	HK	6,720	27,896	**27,906,879**	2,892	16,040	**12,508,796**	8,740	43,936	**40,450,291**
South Africa	ZA	7,318	28,271	**31,683,286**	4,566	16,993	**13,645,623**	10,308	45,264	**45,364,498**
Nigeria	NG	3,448	23,329	**30,622,738**	2,072	13,956	**11,996,583**	4,516	37,285	**42,646,098**
Ghana	GH	3,161	32,189	**27,644,721**	1,053	15,162	**11,088,160**	3,616	47,351	**38,768,231**
Kenya	KE	4,222	31,166	**28,552,920**	2,073	14,796	**12,480,777**	5,193	45,962	**41,069,085**
Tanzania	TZ	3,829	27,533	**24,883,840**	1,414	13,823	**10,253,840**	4,575	41,356	**35,169,042**
Jamaica	JM	3,049	30,928	**28,505,416**	1,049	15,820	**11,124,273**	3,488	46,748	**39,663,666**
TOTAL		**220,405**	**1,140,741**	**1,300,348,146**	**170,224**	**651,304**	**583,923,681**	**340,619**	**1,792,045**	**1,885,632,973**

The Nigerian component of the corpus contains 42.6 million words. This is the largest corpus of Nigerian English. Having such a large corpus of English language as it is being used in Nigeria can provide data on how frequently the collocations the learners produced are used in Nigeria. We may be able understand from the frequency data whether the learners used certain collocations more frequently and correctly because they are frequently used in their

speech community. GloWbE provides frequency data with five different shades of blue with the deepest shade of blue being the highest frequency as can be seen in table 6.1 below.

Table 6. 1 Screenshot of frequency data from GloWbE (Davies, 2013).



	CONTEXT	ALL	US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK	ZA	NG	GH	KE	TZ	JM
1	ECONOMIC RECESSION	1393	190	86	233	92	41	66	85	53	26	79	42	36	41	59	58	51	28	43	21	63
2	ECONOMIC RECESSIONS	108	26	16	14	7	4	4	3	4	2	3	6	2		3	3	3	2	2	2	2
3	ECONOMIC RECESSION/DEPRESSION	4		1	2		1															
4	ECONOMIC RECESSION/TURMOIL	2		2																		
5	ECONOMIC RECESSION/RECOVERY	2	2																			
6	ECONOMIC RECESSIONARY	1							1													
7	ECONOMIC RECESSION/DOWNTURN	1	1																			
8	ECONOMIC RECESSION.THEIR	1				1																
9	ECONOMIC RECESSION.	1															1					
10	ECONOMIC RECESSION--AN	1	1																			
	TOTAL	1514	220	105	249	100	46	70	89	57	28	82	48	38	41	62	62	54	30	45	23	65

1.422 seconds

For ease of analysis, I have coded these shades of blue with numbers one to five with number one being the lowest frequency and five the highest frequency. For example, 'economic + recession' in table 6.1 will be in frequency category 5 for Bangladesh (BD) category 4 for Jamaica (JM); category 3 for Nigeria (NG), category 2 for Ghana (GH) and category 1 for Australia (AU). I will now proceed to investigate the relationship between the collocations produced by the learners and the frequency of those collocations in the Nigerian component of GloWbE.

6. 3 Effects of Frequency of Input on Production of Collocations: Verb Noun Collocations

I will analyse the frequency data in four sub-sections. Firstly, I will analyse the relationship between the frequently used incongruent verb noun collocations in NILECORP-C1 and the frequency data from the Nigerian component of GloWbE. Secondly, the analysis will focus on less frequently used incongruent verb noun collocations in the learner corpus. I will do the same with both frequently used and less frequently used congruent verb noun collocations in the learner corpus in sub-section three and four respectively. I consider any of the verb noun collocations that appear in the NILECORP-C1 four times and below to be less frequently used while the ones that appear five times and above to be frequently used. Similarly, the collocations that fall below category 3 of the frequency data in the Nigerian component of GloWbE is regarded as not frequently used in Nigeria. But the ones that are in category 3 and above are regarded as frequently used.

Before proceeding to the presentation of data and analysis, it would be helpful to be reminded that English is a second language in Nigeria. For some, English is their only language and the other majority, English is their second language. What this means is that, Nigerian learners of English get exposed to the linguistic input beyond the language classroom. So, they learn the language both in the classroom and incidentally outside the classroom. Considering the learners' context, it is highly probable that the learners would be frequently exposed to collocations that are frequently used in the Nigerian component of GloWbE outside the classroom. This possibility will be taken into consideration when interpreting the findings.

6.3.1 Frequently used Incongruent Verb Noun Collocations in NILECORP-C1

There are twenty-one collocational structures in this category. These verb noun collocations are frequently used by the learners. All these verb + noun collocational structures extracted from the learner corpus, which are incongruent, are also present in the Nigerian component of GloWbE. With this, we can use the frequency data to determine whether these structures are

frequently used in the learners' speech community or not. This data can help to draw an inference on the effects of frequency and possible exposure to the input on the learners' ability to produce incongruently verb noun collocations accurately. Table 6. 2 presents the collocations and the frequency data for both corpora.

Table 6.2 Frequency data from NILECORP-C1 and GloWbE

Frequently used incongruent Verb Noun Collocations			
Collocations	**Frequency in NILECORP-C1**	**Frequency Category in GloWbE**	**Frequency in GloWbE**
express + view	46	3	156
give + birth	42	5	956
take + care	39	5	2,660
make + friend	30	4	339
make + money	20	5	5,868
keep + gang	10	1	0
take + bath	10	5	72
write + examination	9	5	167
tarnish + image	9	5	100
bear + children	9	4	131
perform + task	9	3	68
commit + crime	8	5	327
take + decision	8	5	477
handle + equipment	7	1	13
fall in + love	7	4	542
damage + image	6	1	9
heed + advice	6	5	54
pamper + children	5	1	3

achieve + goal	5	5	627
pay + attention	5	5	869
narrate + story	5	3	44
Total usage	**295**		

There are 21 frequently used incongruent verb noun collocational structures in this table.

Seventeen out of the twenty-one collocational structures representing 80.9% in the category are frequently used in GloWbE. They are in the top three frequency categories (the three deepest shades of blue of the frequency data). The other four structures representing 19.1% are not frequently used in the corpus. They are in the lower categories of the frequency table. These structures are: 'keep + gang' which was used ten times by the learners, 'handle + equipment' was used seven times by the learners, 'damage + image' was used six times by the learners, and 'pamper + children' which was used five times by the learners. All these structures were correctly produced by the learners and used frequently in the learner corpus even though they have no equivalent in their L1. The structures: 'handle + equipment' and 'damage + image' have elements of idiomaticity which might make the processing considerably challenging for learners. Besides, these collocational structures are not frequent expression in Nigeria as suggested by the frequency data from the Nigerian component of GloWbE.

In summary, a clear majority (80.9%) of the incongruent verb noun collocations frequently used by the learners are also very frequently used in the Nigerian component of GloWbE. Incongruent collocations are said to be problematic for learners but the learners in this study have produced far more than expected incongruent collocations. Could it be the case that frequency (and exposure) of input trumps incongruency? A Spearman's correlation was run to determine the relationship between the 21 frequently used incongruent verb noun collocational structures in NILECORP-C1 and the same collocational structures in the

Nigerian component of GloWbE. There was a medium correlation between them (r_s = .31, n = 21, p < 0.15). However, it is not statistically significant.

6.3.2 Less Frequently used Incongruent Verb Noun Collocations in NILECORP-C1

This sub-section investigates the relationship between the frequency of the target structure in the Nigerian component of GloWbE and the production of less frequently used collocation in the learner corpus. See table 6.3 below for more details:

Table 6.3 Frequency data for less frequently used incongruent verb noun collocation in NILECORP-C1

Less Frequently used Incongruent Verb Noun Collocations			
Collocations	**Frequency in NILECORP-C1**	**Frequency Categories in GloWbE**	**Frequency in GloWbE**
fetch + water	4	4	106
acquire + knowledge	4	4	104
impart + knowledge	3	3	32
form + habit	3	5	44
contract +disease	3	1	22
achieve + dream	3	3	130
say + prayer	3	3	84
make + love	3	5	260
lavish + money	2	1	8
keeping + late + night	2	1	1
draw + attention	2	5	438
catch + attention	2	3	120
turn down + offer	2	3	24

keep + record	2	4	120
hold + position	2	5	104
shed + blood	2	5	115
shed + tears	2	4	152
obtain + money	2	5	45
break + regulation	2	1	0
tarnish + reputation	2	3	29
Total usage	**50**		

The frequency data from the Nigerian component of GloWbE indicates that sixteen (representing 80%) out of the twenty structures in this category are frequently used in Nigeria. The four structures which are not frequently used in the Nigerian component are: 'break + regulation', 'keeping + late night', 'lavish + money', and 'contract + disease'. The structure 'contract + disease' was used in the learner corpus three times while the others were used twice each. There seems to be a pattern of the learners using more frequently the collocations that are frequently used in the Nigerian component of GloWbE. But there are also instances, although few, of the learners producing incongruent collocations that are not frequently used in the Nigerian component of GloWbE. A Spearman's correlation was run to determine the relationship between the 20 less frequently incongruent used verb noun collocational structures in NILECORP-C1 and the same collocational structures in the Nigerian component of GloWbE. There was a small correlation between them (r_s = .11, n = 20, p < 0.64) but not statistically significant.

To conclude this section on the effect of frequency and possible exposure to input on the production of incongruent verb noun collocations, the data reveals that thirty three out of the forty-one incongruent verb noun collocational structures (both frequently used and not frequently used) the learners produced correctly representing 80.48% are frequently used in Nigeria. Only eight structures representing 19.52% are not frequently used in the Nigerian component of GloWbE.

6.3.3 Frequently used Congruent Verb Noun Collocations in NILECORP-C1

This sub-section analyses the relationship between the frequency data in GloWbE and the frequently used congruent verb noun collocations in the learner corpus. There are thirty collocational structures in this category. These thirty congruent verb noun collocational structures were collectively produced three hundred and nineteen times.

Table 6. 4 Frequently used congruent collocations

Frequently used congruent Verb Noun Collocations					
Collocations	**Frequency in NILECORP**	**Frequency in GloWbE**	**Collocations**	**Frequency in NILECORP**	**Frequency in GloWbE**
lead + trouble	40	5	drive + car	7	4
have + sex	27	5	have + experience	7	5
take + hospital	26	5	cause + trouble	7	1
listen + advice	21	5	make + trouble	7	1
write + story	19	4	fail + examination	6	1
cause + problem	16	1	have + accident	6	5
tell + story	13	5	bring + shame	5	5
rush + hospital	12	5	learn + lesson	5	5
have + dream	11	5	make + decision	5	5
pass + examination	9	3	ask + forgiveness	5	3
spend + money	8	5	have + tendency	5	3
have + problem	8	5	have + opportunity	5	4
go + hospital	8	5	keep + secret	5	4
give + treatment	8	4	look + trouble	5	5
take + decision	8	5	pick + phone	5	5

As is the case with incongruent verb noun collocations, the learners tend to use congruent verb noun collocations that are commonly used in their speech community. All the thirty structures apart from four are frequently used in Nigeria. The four most frequently used

congruent verb noun collocations in the learner corpus are also highly frequently used in the GloWbE. The structures which are not frequently used the Nigerian component of GloWbE are: 'cause + problem', 'cause + trouble', 'make + trouble' and 'fail + examination'. But again, over 86% of all the collocations the learners produced in this category are frequently used in GloWbE. This points to a link between frequency of input/exposure to input and production of collocations.

6.3.4 Less frequently used Congruent Verb Noun Collocations in NILECORP-C1

The collocations in this data subset appear between two and four times in the learner corpus. There is no evidence to suggest that the learners used these congruent collocations less frequently in their written text because they have difficulty producing them. On the contrary, the learners' production of fewer of these collocations may be down to the communicative needs and the genre of the written text.

Table 6.5 Less frequently used congruent collocations

Less frequently used congruent Verb Noun Collocations		
Collocations	**Frequency in NILECORP**	**Frequency in GloWbE**
pay + money	4	5
give + instruction	4	5
collect + result	4	1
follow + rule	4	3
burst + tears	4	3
accept + offer	3	5
have + friend	3	5
read + story	3	5

follow + instruction	2	5
send + prison	2	1
give + punishment	2	1
Total Usage	**35**	

Eight out of the eleven collocational structures in this category appear frequently in the Nigerian component of GloWbE. This is 72.7% of all the collocations in the category. A pattern can be identified in the relationship between the frequency of the collocations in the Nigerian component of GloWbE and the verb noun collocations produced by the learners. This pattern is the same with incongruent and congruent verb noun collocations. Thirty-four representing 82.9% out of the forty-one congruent verb noun collocations produced by the learners in their written text appear frequently in the Nigerian component of GloWbE. Only seven representing 17.9% were not frequently used in GloWbE.

6. 4 Effects of Frequency of Input on Production of Collocations: Adjective Noun Collocations

Using the same approach and procedure I used to analyse the verb noun collocations, this section is also divided into four sub-sections. Firstly, I will analyse the relationship between the frequently used incongruent adjective noun collocations in NILECORP-C1 and the frequency data from the Nigerian component of GloWbE. Secondly, the analysis will focus on less frequently used incongruent adjective noun collocations in the learner corpus. I will do the same with both frequently used and less frequently used congruent adjective noun collocations in the learner corpus in sub-section three and four respectively. Any of the adjective noun collocations that appear in the NILECORP-C1 four times and below are considered be less frequently used while the ones that appear five times and above are frequently used. Similarly, the collocations that fall below category 3 of the frequency data

in the Nigerian component of GloWbE is regarded as not frequently used in Nigeria while the ones that are in category 3 and above are regarded as frequently used.

6.4.1 Frequently used Incongruent Adjective Noun Collocations in NILECORP-C1

According the data on this sub-section (see table 6.1 below), the effect of frequency on the production of collocations seems more pronounced on the most frequently used adjective noun collocations by the learners. 90% of the collocations they produced in this category are also frequently used in the Nigerian component of GloWbE. The only combination out of the ten produced which is not frequently used in GloWbE is 'bad + disposition' but which the learners used 11 times. A search on how this combination is used across the twenty countries in the GloWbE corpus indicates that the combination is sparingly used in English. It is not clear why the learners have used it frequently in their texts despite it being less frequently used in Nigerian English. However, as can been seen in the previous analysis, there is a relationship between frequency of input and collocation production.

Table 6.1 Frequently used Incongruent Adjective Noun Collocations in NILECORP-C1

Frequently used incongruent Adjective Noun Collocations			
Collocations	**Frequency**	**Frequency Categories in GloWbE**	**Frequency in GloWbE**
peer + group	28	3	38
secret + cult	28	5	93
bright + future	13	3	120
elder + brother	12	5	246
long + time	11	3	2747
bouncing + baby	11	5	68

bad + disposition	11	1	105
unwanted + pregnancy	10	5	406
nearest + future	5	5	69
major + problem	5	5	0
Total Usage	**134**		

There are 10 frequently used incongruent adjective noun collocational structures in this table.

A Spearman's correlation was run to determine the relationship between the 10 frequently used incongruent adjective noun collocational structures in NILECORP-C1 and the same collocational structures in the Nigerian component of GloWbE. There was a negative correlation between them (r_s = -0.21, n = 10, p < 0.54) though it is not statistically significant.

6.4.2 Less Frequently used Incongruent Adjective Noun Collocations in NILECORP-C1

This sub-section is essentially aimed at finding out whether these collocations were used fewer times because they were less frequently used in Nigerian English. There are thirteen collocational structures in this category. But as it turned out 76.9% of these collocations are frequently used in the Nigerian component of GloWbE. One of the three collocational expressions which was not frequently used in GloWbE is 'tight + friend'. It is used to refer to intimate/close friendship in Nigerian English, but it is sometimes frowned on by people who want to strictly adhere to British English norms. Despite this, it is a common expression in social discourse and particularly, among the older generations. See table 6.2 for more details on the frequency data.

Table 6.2 Less frequently used Incongruent Adjective Noun Collocations in NILECORP-C1

Less frequently used incongruent Verb Noun Collocations			
Collocations	**Frequency**	**Frequency Categories in GloWbE**	**Frequency in GloWbE**
last + night	4	2	10
social + miscreant	4	4	32
peer + pressure	4	3	15
average + student	3	3	35
curious + mind	3	5	4
nonchalant + attitude	2	5	15
deep + voice	2	4	31
tight + security	2	4	26
strong + influence	2	3	48
pipe-borne + water	2	3	111
varying +degree	2	5	791
tight + friend	2	1	1
bright + student	2	1	9
Total Usage	**34**		

There are 13 less frequently used adjective noun collocational structures in this table.

Apart from the frequently used incongruent adjective noun collocations, there seems to be trend in this data which suggests a positive relationship between frequency of input in the learners' speech community and collocation production. This has been the case with both incongruent and congruent verb noun collocations as well as incongruent adjective noun collocations. A Spearman's correlation was also run to determine the relationship between

the 13 less frequently used incongruent adjective noun collocational structures in NILECORP-C1 and the same collocational structures in the Nigerian component of GloWbE. There was a very weak correlation between them (r_s = 0.09, n = 13, p < 0.75). Again, like the other Spearman's correlation test, this too is not statistically significant. I will now analyse the congruent adjective noun collocations.

6.4.3 Frequently used Congruent Adjective Noun Collocations in NILECORP-C1

This data sub-set reveals that 80% of the collocations are frequently used in the Nigerian component of GloWbE which is consistent the findings in the previous sub-sections. Only five collocational combinations are not frequently used in the GloWbE. These combinations are: 'fellow + student', 'moral + education', 'final + examination' 'bad + language' and 'bad + influence'. 'Bad + influence' and 'fellow + student' were used very frequently in the learner corpus probably because of the theme the learners were writing about.

Table 6.3 Frequently used Congruent Adjective Noun Collocations in NILECORP-C1

Frequently used congruent Adjective Noun Collocations		
Collocations	**Frequency in NILECORP**	**Frequency in GloWbE**
good + friend	56	4
bad + behaviour	40	5
bad + character	30	5
best + friend	27	3
fellow + student	23	1
bad + attitude	17	3
bad + influence	15	2
armed + robbery	14	5

bad habit	10	3
good + manner	9	3
final + examination	8	2
class + mate	7	3
best + student	7	5
moral + education	6	1
bad + language	6	2
good + care	6	5
close + friend	5	5
second + wife	5	3
good + news	5	4
main + purpose	5	4
big + shame	5	5
import + position	5	5
bad + manner	5	4
bad + reputation	5	4
good + record	5	5
Total Usage	**326**	

6.4.4 Less Frequently used Congruent Adjective Noun Collocations in NILECORP-C1

The last of these data sub-sets reveal similar findings. Over 60% of the collocations produced by the learners are also frequently used in the Nigerian component of GloWbE. Only four out of the twelve collocational structures produced by the learners are not frequently used in GloWbE. These findings are also consistent with the previous ones. See table 6.4 for more details.

Table 6.4 Less Frequently used Congruent Adjective Noun Collocations in NILECORP-C1

Less Frequently used Congruent Adjective Noun Collocations		
Collocations	**Frequency**	**Frequency in GloWbE**
little + brother	4	2
hard + work	4	5
good + character	4	4
true + friend	3	5
age + group	3	2
several + times	3	5
innocent + blood	3	5
secret + society	2	4
bad + news	2	4
evil + companion	2	4
quick + reaction	2	2
sweet + experience	2	1
Total Usage	**34**	

6.5 Summary of Findings

The data clearly reveals that 52 representing 81.2% out of the 64 incongruent collocational structures extracted from the learners' texts are frequently used in the Nigerian component of GloWbE. And 62 representing 78.4% out of the 78 congruent collocational structures produced by the learners are frequently used in GloWbE. Overall, 80.2% of all the collocational structures produced by the learners (114 out of 142) are frequently used in the Nigerian component of GloWbE. The data also reveals the learners produced 28

collocational structures representing 19.8% of all the structures produced that are not frequently used in the Nigerian component of GloWbE. Of these figures, 12 are incongruent while 16 are congruent. The findings suggest that frequency of input and apparent exposure to the input outside the classroom facilitate the production of collocations. It also suggests that the production of collocations is not entirely the function of frequent exposure to the input. While incongruent collocations are said to be problematic for learners, frequently used incongruent collocation in the learners' speech community, as this data reveals, seem to be less problematic. I will now interpret and explain these findings in relation to the immediate literature

6.6 Discussion

Three themes imaged from the findings: (1) that frequency and exposure to input facilitate the productive knowledge of collocations, (2) that production of collocation is not entirely the function of frequent exposure to input (3) that frequency trumps incongruency. Starting with the first theme, considering the trend in the findings, there is a strong evidence to conclude that L2 learners acquire more of the collocations that are frequently used in their speech community. This corroborates González Fernández and Schmitt's (2015) findings that learners' knowledge of collocations correlates moderately with corpus frequency and everyday engagement with English outside the classroom This seems to support the idea that usage-based models of language acquisition can be fruitfully applied to understanding the processes that underlie L2 collocational acquisition (Wolter & Gyllstad, 2013). So, frequency of the target structure in the input seems to be key to the acquisition of collocations. This highlights the role of immersion-based L2 exposure in collocational development and competence. In the case of Nigeria where these learners live, English is a second language. This context provides an immersion-based environment where they will frequently encounter the target structure in various settings. A number of studies have investigated the effect of immersion on the production of L2 formulaic sequences (Nesselhauf, 2005; Waibel, 2008; Siyanova & Schmitt, 2008; Groom, 2009). But I will focus on Nesselhauf (2015 and Groom (2009), two studies with opposing findings, to explain this.

Nesselhauf investigates the effect of immersion on the acquisition of collocations from a phraseological perspective. Her analysis of the German learner corpus of English (GeCLE) reveals, "the length of stays in English speaking country does not seem to lead to an increased use of collocations; instead, there even seems to be a slight trend in the opposite direction" (Nesselhauf, 2005: 236). Her findings are obviously counterintuitive because of the widely held assumption that the best way to learn a language is to live in the target language context. But if we consider her findings in the light of the fact that 19.2% of the collocations the learners in my study produced are not frequently used in the Nigerian component of GloWbE, it would suggest that acquisition of collocation is not entirely the function of frequency and exposure to the input. Having said that, it is important to have a caveat here. It is difficult to determine individual learner's exposure to input. Beyond what the corpus frequency data suggests, we cannot be very sure of what input learners are exposed to in their personal language acquisition experience. But if the structures are frequent in the input the learners are exposed to in the immersion environment, it is plausible to approximate the level of exposure the learners might have. Even then this is a slippery ground because the next question that would come to mind is: Does frequent exposure to the input mean learners will always notice the target structure?

Groom's (209) study on the other hand, which was a response to Nesselhauf's study on the effect of immersion on the acquisition of collocations lends credence to the role of frequency and exposure to input on the acquisition of collocations. He analyses a similar corpus albeit from the frequency-based perceptive and comes up with a different conclusion. He uses the Uppsala Student English corpus (USE), a bigger corpus than the GeCLE used by Nesselhauf. Using both lexical bundle analysis and node and collocate analysis approaches, he analyses the text of two groups of Swedish learners of English – Immersion and Non-immersion. While Nesselhauf (2005) does not see any significant correlation between the time the learners spent in L2 context and their collocational competence, Groom (2009: 33) discovers that "not only that collocational accuracy does appear to be more positively correlated with L2 immersion, but also that the difference between immersion and non-immersion group may be more substantial than Nesselhauf (2005) suggests". What this suggests is that learners living or learning the target structure in the L2 context have a greater chance of exposure to the input frequently in and outside the classroom. This facilitates the production of collocations. The fact that over 80% of the collocations the learners in this study produced are

frequently used in their speech community is an evidence of the positive effect of frequency (and exposure) of input on the acquisition of collocations.

If we conclude that frequency of instances of collocations in the input the learners are exposed to in some ways facilitates acquisition, does that suggest that learners would not have difficulty producing frequent collocation? We could have answers to this question in the analysis of the collocational errors produced by these learners. We will be able to find out if the problematic collocations are frequent in Nigerian speech community or not. But before the error analysis, the next chapter will focus on the relationship between the production of collocations and proficiency.

Chapter Seven

Production and Use of Collocations across Proficiency Levels

7.0 Introduction

This chapter enquires into the relationship between language proficiency and the production of verb noun and adjective noun collocations by L2 learners. The last two chapters have revealed that the difference in the collocations produced by the learners and native speakers lies mainly in their linguistic complexity in terms of their collocation span and the structural and semantic properties of their constituents; and that as input increases collocational output also increases. This chapter, therefore, attempts to find out if L2 learner's knowledge of collocations increases in tandem with their general proficiency in the English language. It considers the following research questions: (1) What is the relationship between proficiency and the production of verb noun and adjective noun collocations? (2) What is the relationship between proficiency and the production of incongruent verb noun and adjective noun collocations? (3) What is the relationship between proficiency and the production of congruent verb noun and adjective noun collocations? (4) What is the relationship between proficiency and the use of linguistically complex verb noun collocations in terms of the collocation span and the structural properties of their constituents? (5) What is the relationship between proficiency and the use collocations with additional nuances and associations – the degree of semantic opacity and transparency? As part of this investigation, I will analyse the verb noun and adjective noun collocations produced by four groups of Yoruba-speaking English learners representing four proficiency levels which are equivalent to the Common European Framework of Reference for Language C1, B2, B1 and A2.

As stated earlier, according to Council of Europe (2001), L2 learners at C1 proficiency level of the Common European Framework of Reference for Language can "*understand* a *wide range* of *demanding*, *longer texts*, and recognise *implicit meaning" (CoE, 2001: 24)*. Learners at this stage can express themselves fluently and can use language spontaneously without struggling to find expressions. They can produce *"well-structured, detailed text* on *complex* subjects, *showing controlled use of organisational patterns" (CoE, 2001: 24)* as

well as cohesive devices and achieve cohesion in their expressions either written or spoken. Learners at the B2 proficiency level can *"understand the main ideas* of *complex text* on both concrete and ... *degree* of *fluency* and *spontaneity* that makes regular *interaction* with native speaker" (CoE, 2001: 24). Just like the C1 level, Learners at the B2 level can produce clear and detailed text on wide range of topics. Learners at the B1 proficiency level can comprehend main ideas of clear standard input on issues they frequently encounter in their environment. They are proficient enough to produce simple connected texts on issues which are familiar to them. And finally, learners at A2 proficiency level can comprehend structure and frequently used expressions related to their areas of relevance. They can communicate in simple and routine task.

This chapter is divided into seven sections as follows:

The first section presents the overall descriptive statistics of the data used for this study. It includes numbers of tokens in the four sub-corpora, the numbers of verb noun and adjective noun collocations extracted from the corpora, the congruent and incongruent collocations produced, and the semantically burdensome collocations produced by each of the four proficiency groups. In the second section, the analysis will focus on identifying, comparing and interpreting evidence from the four sub-corpora. This is to determine the difference in the overall collocations production across the four proficiency levels. Sections three and four will focus on fine-grained analysis of the collocations produced to determine how many of them are incongruent and how many are congruent. This will show the relationship between proficiency and the production of incongruent and congruent verb noun and adjective noun collocations. Section five presents a qualitative analysis of the linguistic complexity of the verb noun collocations produced by each proficiency level. This analysis will focus on the span of the collocations and the structural properties of their constituents. This section is divided into two sub-sections: one focuses on collocation span while the other focuses on the structural properties of their constituents. In section six, the data on collocations with modified meanings to introduce additional nuances and associations (Phillip, 2011) will be analysed across the proficiency levels. These collocations, as I have stated earlier in chapter five, have meanings beyond the surface meaning of the lexical items constituting the collocations. The aim of the analysis in this section is to determine the relationship between the use of such collocations and proficiency. Finally, in the seventh section, which is a

discussion section, I will interpret and explain my findings and examine whether and how my research questions have been answered. The discussion will show how my findings relate to the immediate literature on the relationship between proficiency and L2 learners' use of collocations.

7. 1 Overall Results

The descriptive statistics presented here describe the basic features of the data used to investigate the relationship between proficiency and L2 learners' production and use of collocations. It provides simple summaries about the samples and measures used in this section. Four sub-corpora were used in this study – NILECORP-C1, NILECORP-B2, NILECORP-B1 and NILECORP-A2. NILECORP-C1, the most proficient group of the four learner groups has 252,003 word tokens and 9,193 word types. NILECORP-B2 has 130,559 word tokens and 6,322 word types. NILECORP-B1 has 73,660 word tokens and 2,197 word types while NILECORP-A2, the least proficient group has 66,996 word tokens and 4,555 word types. All the nouns involved in the study appear, at least, six times in the corpora and only verb + noun and adjective noun collocations that occur twice and above were included in the analysis. All nouns that appear fewer than six times and all instances of verb noun and adjective noun collocations that appear fewer than two times were excluded from the analysis.

A total of 2,397 collocations were extracted from the Nigerian Learner Corpus – 1,324 from NILECORP-C1, 599 from NILECORP-B2, 213 from NILECORP-B1 and 261 from NILECORP-A2. Out of the 1, 324 collocations produced by the NILECORP-C1 group, 793 are verb noun collocations while 531 are adjective noun collocations. Three hundred and seventy seven of the 599 collocations produced by the NILECORP-B2 group are verb noun collocations while 222 are adjective noun collocations. The NILECORP-B1 group produced 164 verb noun collocations and 49 adjective noun collocations while the NILECORP-A2 group produced 234 verb noun collocations and 27 adjective noun collocations. See table 7.1 for more details:

Table 7.1 Proficiency Groups Statistics

Proficiency Groups	Corpus Size	Verb Noun Collocations	Adj. Noun Collocations	Total Collocations
NILECORP-C1	252,003	793	531	1,324
NILECORP-B2	130,559	377	222	599
NILECORP-B1	73,660	164	49	213
NILECORP-A2	66,996	234	27	261
Total	**523,218**	**1,568**	**829**	**2,397**

All the four groups of learners produced both congruent and incongruent collocations apart from the NILECORP-A2 group – the least proficient group – which did not produce incongruent adjective noun collocations. They only produced adjective noun collocations that have their L1 equivalent. See tables 7.2 and 7.3 below for more details:

Table 7.2 Overall Verb Noun Collocations

Overall Verb Noun Collocations				
Proficiency Groups	Corpus Size	Verb Noun Collocations	Incongruent Verb Noun Collocations	Congruent Verb Noun Collocations
NILECORP-C1	252,003	793	345	448
NILECORP-B2	130,559	377	125	254
NILECORP-B1	73,660	164	102	62
NILECORP-A2	66,996	234	70	164

Table 7.3 Overall Adjective Noun Collocations

Overall Adjective Noun Collocations				
Proficiency Groups	Corpus Size	Adj. Noun Collocations	Incongruent Adj. Noun Collocations	Congruent Adj. Noun Collocations
NILECORP-C1	252,003	531	144	387
NILECORP-B2	130,559	222	23	199
NILECORP-B1	73,660	49	6	43
NILECORP-A2	66,996	27	0	27

In the next section, I will attempt to interpret the above statistics to show what it means in relation to the four proficiency levels and their production of collocations.

7.2 Collocation Production across Four Proficiency Levels

As shown above the NILECORP-C1 group produced a total of 1,324 verb noun and adjective noun collocations. If we consider the corpus size of 252,003 words, this would be 0.52% (number of collocations divided by the size of corpus multiplied by hundred – 1,324 ÷ 252,003 × 100). The second group, NILECORP-B2, produced 599 collocations. Again, if we consider this in relation to the corpus size of 130,559 words, that would be 0.45%. The NILECORP-B1 group, which is the second least proficient group, produced 213 collocations. If we consider the corpus size of 73,660 words, this amounts to 0.28%. Up to this point, we can see consistent progression in the production of collocations from the least proficient to the most proficient group of learners. However, the NILECORP-A2 group which is the least proficient of the four learner groups produced more collocations than the NILECORP-B1 group. They produced 261 collocations. Considering the size of the sub-corpus of 66,996 words, that is 0.38%. The analysis at this stage focuses only on the number of instances of verb noun and adjective noun collocations in each of the sub-corpus. So, based on the frequency of the collocations used by each group regardless of the number of times certain collocational structures were repeated, data from C1, B2 and B1 proficiency levels suggest

the higher the proficiency, the more collocations L2 learners produce. See table 7.4 below for more details:

Table 7.4 Collocations Produced Across the four Learner Groups

Proficiency Groups	Corpus Size	Verb Noun Collocations	Adj. Noun Collocations	Total Collocations	Percentage
NILECORP-C1	252,003	793	531	1,324	0.52%
NILECORP-B2	130,559	377	222	599	0.45%
NILECORP-B1	73,660	164	49	213	0.28%
NILECORP-A2	66,996	234	27	261	0.38%

Meanwhile, the A2 group will be investigated further to determine why they seem to produce a result that is inconsistent with the other learner groups. The least proficient group producing more collocations than the second least proficient group throws up some questions which need to be investigated further. With this, the next phase of the investigation focuses on the number of collocational structures produced by each group as opposed to the overall frequency of the instances of collocations. Each different collocational structure is counted once regardless of how many times it was used in the text. As a reminder, collocational structures like 'commit a heinous crime' and 'commit a serious crime' are regarded as the same 'commit + crime' structure whereas 'convicted of a heinous crime' and 'commit a heinous crime' are two difference collocational structure – 'commit + crime' and 'convict + crime'.

As shown in table 7.5 below, the NILECORP-C1 group produced 148 collocational structures, NILECORP-B2 produced 92, NILECORP-B1 produced 39 while NILECORP-A2 produced 28 structures.

Table 7.5 Collocational Structures across the four Learner Groups

Proficiency Groups	Corpus Size	Verb Noun Collocational Structures	Adj. Noun Collocational Structures	Total Collocational Structure	Percentage
NILECORP-C1	252,003	88	60	148	0.058%
NILECORP-B2	130,559	73	19	92	0.070%
NILECORP-B1	73,660	29	10	39	0.052%
NILECORP-A2	66,996	23	4	27	0.040%

Considering the numbers of collocational structures produced in proportion to the corpus sizes, NILECORP-C1 is 0.058%, NILECORP-B2 is 0.070%, NILECORP-B1 is 0.052% while NILECORP-A2 is 0.041%. Again, the result is consistent with the earlier findings apart from the B2 proficiency group which appears to produce more collocational structures than the C1 group which is the most proficient group. What this means is that, in quantitative terms, there seems to be a convincing progression in the production of collocations as learners' proficiency increases. Having said that, the apparent inconsistency in the result for A2 group's production of collocations and B2 group's production of collocational structures suggests that despite rigorous attempt to assign proficiency to these groups of learners, there may still be some outliers. The inconsistency in the result could be ascribable to a few learners in the groups who could be more proficient than the rest of the groups or who may have more exposure to certain collocations outside the classroom. This seems be an isolated case as only one out of four learner groups in each category was affected. Meanwhile, the value and difficulty of assigning proficiency to corpus texts will be addressed in chapter nine. From the next section, I will do fine-grained analysis of the collocations production by each proficiency group to have a comprehensive understanding of the relationship between proficiency and collocation production.

7.3 Production of Incongruent and Congruent Verb Noun Collocations across Proficiency Levels

All the four proficiency groups except group B1 produced more congruent verb noun collocations than incongruent collocations. The result for groups C1, B2 and B1 is consistent with other findings in the literature which suggest congruent collocations are easier for learners to produce (Peters, 2016). See table 7.6 below for more details, and the incongruent verb noun collocations produced by groups C1, B2, B1 and A2 are in Tables 7.7, 7.8, 7. 9 and 7.10. Meanwhile, I am representing the incongruent verb noun collocations produced by the C1 group in table 7.7 for ease of analysis and comprehension.

Table 7.6 Overall Verb Noun Collocations

Overall Verb Noun Collocations				
Proficiency Groups	Corpus Size	Verb Noun Collocations	Incongruent Verb Noun Collocations	Congruent Verb Noun Collocations
NILECORP-C1	252,003	793	345	448
NILECORP-B2	130,559	377	125	254
NILECORP-B1	73,660	164	102	62
NILECORP-A2	66,996	234	70	164

Table 7.7 Incongruent Verb Noun Collocations Produced by C1 Group

Incongruent Verb Noun Collocations Produced by C1 Group			
	Frequency		Frequency
express + view	46	narrate + story	5
give + birth	42	fetch + water	4
take + care	39	acquire + knowledge	4
make + friend	30	impart + knowledge	3
make + money	20	form + habit	3

keep + gang	10	contract +disease	3
take + bath	10	achieve + dream	3
write + examination	9	say + prayer	3
tarnish + image	9	make + love	3
bear + children	9	sentence + prison	3
perform + task	9	keeping + late + night	2
commit + crime	8	draw + attention	2
take + decision	8	catch + attention	2
handle + equipment	7	turn down + offer	2
fall in + love	7	keep + record	2
damage + image	6	lavish + money	2
heed + advice	6	obtain + money	2
pay + attention	6	break + regulation	2
achieve + goal	5	tarnish + reputation	2
pamper + children	5	engage + prostitution	2
		Total	345

The table contains 40 different verb noun collocation structures which were used 345 times in the corpus.

Table 7.8 Incongruent Verb Noun Collocations Produced by B2 Learners

Incongruent Verb Noun Collocations Produced by B2 Learners			
Collocations	Frequency	Collocations	Frequency
rush + hospital	11	commit + abortion	3
make + friend	8	regain + consciousness	3
donate + kidney	8	perform + surgery	2
prescribe + drug	8	save + money	2
fetch + water	7	raise + money	2

donate + blood	6	pay + attention	2
acquire + knowledge	6	perform + operation	2
take + bath	6	carry out + operation	2
prescribe + medicine	5	gain + admission	2
write + examination	4	share + knowledge	2
risk + life	4	achieve + dream	2
pass on + knowledge	3	harvest + crop	2
take care + patient	3	sit + examination	2
spend + time	3	watch + movie	2
sentence + death	3	make + difference	2
share + problem	3		
perform + task	3	Total	**123**

The table contains 32 different incongruent verb noun collocational structures which were used 123 times in the corpus.

Table 7.9 Incongruent Verb Noun Collocations Produced by B1 Learners

Incongruent Verb Noun Collocations Produced by B1 Learners			
Collocations	Frequency	Collocations	Frequency
give + birth	43	have + breakfast	3
take + bath	11	watch + film	3
snap + picture	8	take + picture	2
brush + teeth	6	fetch + water	2
ride + horse	6	embark + journey	2
spend + holiday	5	ride + bicycle	2
go on + holiday	5		
take + breakfast	4	**Total**	**102**

There are 14 different Incongruent Verb Noun Collocational Structures in this table; they were used 102 times.

Table 7.10 Incongruent Verb Noun Collocations Produced by A2 Learners

Incongruent Verb Noun Collocations Produced by A2 Learners			
Collocations	Frequency	Collocations	Frequency
take + bath	34	take + notice	5
make + friend	18	have + bath	3
give + birth	10	**Total**	**70**

This table contains 5 different incongruent verb noun collocational structures which were collectively used 70 times.

The focus at this stage is to analyse the proportion of the verb noun collocation produced by the learners which are incongruent and how this reflects across proficiency levels. The data shows that 43.3% of the verb noun collocations produced by C1 group are incongruent. The B2 group produced 33.1% incongruent verb noun collocations, B1 produced 62.1% while A2, the least proficient produced 29.1%. Apart from B1 group which produced a different result, all the other three proficiency groups consistently show a progression in the production of incongruent verb noun collocation as their proficiency increases. Three collocational structures in NILECORP-B1 were apparently overused. The structure 'give + birth' was used 43 times, 'go to + church' was used 13 times while 'take + bath' was used 11 times. Using these structures that many times are disproportionate considering the size of the corpus (relatively small in comparison to NILECORP-C1 and NILECORP-B2). And the result for the B1 group which is out of step with the others might be attributable to overuse of these structures. If these three structures were not overused, all the four proficiency groups might have consistently shown a progression in the production of verb noun collocations in tandem with proficiency increase.

This investigation went further to analyse the number of verb noun collocational structures produced by each proficiency group to see the relationship between proficiency and the production of collocational structures but came up with a mixed result. The C1 group

produced 40 verb noun collocational structures, B2 produced 32, B1 produced 14 while A2 produced 5 structures. Considering these in proportion to the size of each sub-corpus by dividing the number of structures by the size of corpus multiplied by 100, C1 is 0.016%, B2 is 0.024%, B1 is 0.019% and A2 is 0.007%. While the data shows that the least proficient group produced the fewest of verb noun collocational structures, it does not however show any consistent progression across the other three proficiency groups. We will now consider the congruent verb noun collocations.

Analysis of the congruent verb noun collocations in relation to the four proficiency groups reveals something that is opposite to what the analysis of the incongruent verb noun collocations suggests. While the production of incongruent verb noun collocations increases as proficiency increases, the production of congruent verb noun collocations decreases as proficiency increases. Starting with the least proficient group, 70.9% of the verb noun collocations produced by the A2 group are congruent, B1 produced 37.9%, B2 produced 66.9% while C1 group produced 56%. Apart from the B1 group, the data suggests the more proficient the learners become, the fewer congruent verb noun collocations they produce. This may mean that as L2 learners become more proficient, they rely less on their L1 to produce L2 structures. This is consistent with Jiang's (2000) model of lexical acquisition which I have discussed in chapter five. Hence, their production of collocations which have no L1 equivalent increases in tandem with proficiency increase while their production of collocations which have L1 equivalent decreases as their proficiency increases. See tables 7.11, 7.12, 7.13 and 7.14 for the congruent verb noun collocations produced by proficiency groups C1, B2, B1 and A2. Meanwhile, I am representing the congruent verb noun collocations produced by the C1 group here for ease of analysis and comprehension.

Table 7.11 Congruent Verb Noun Collocations Produced by C1 Learners

Congruent Verb Noun Collocations			
	Frequency		Frequency
have + friend	45	fail + examination	6
lead to + trouble	40	learn + lesson	5
have + sex	27	ask + forgiveness	5
take + hospital	26	have + tendency	5
listen + advice	21	have + opportunity	5
write + story	19	keep + secret	5
choose + friend	18	look + trouble	5
cause + problem	16	pick + phone	5
spend + money	15	Resume + school	5
tell + story	13	pay + money	4
rush + hospital	12	collect + result	4
have + accident	11	burst + tears	4
have + dream	11	follow + rule	3
have + experience	11	accept + offer	3
pass + examination	9	give + instruction	3
have + problem	8	Collect + result	3
go + hospital	8	shed + tears	3
give + treatment	8	Sentence + prison	3
drive + car	7	hold + position	2
cause + trouble	7	shed + blood	2
make + trouble	7	Fight + war	2
bring + shame	7	follow + instruction	2
make + decision	7	send + prison	2
read + story	7	give + punishment	2
		Total	448

There are a total of 48 different verb noun collocations structures in this table. They were used 448 times in the corpus.

Table 7.12 Congruent Verb Noun Collocations Produced by B2 Learners

Congruent Verb Noun Collocations Produced by B2 Learners			
Collocations	Frequency	Collocations	Frequency
support + motion	36	pour + water	4
save + life	34	have + misunderstanding	4
take + hospital	13	get + admission	4
give + advice	10	make + mistake	3
have + problem	10	have + opportunity	3
have + argument	10	use + opportunity	3
listen + music	9	lose + blood	3
have + accident	8	give + information	3
give + treatment	7	cure + sickness	3
cure + ailment	7	have + knowledge	3
tell + story	6	start + argument	3
oppose + motion	6	answer + question	3
spend + money	5	disturb + peace	2
give + instruction	5	disagree + motion	2
shed + tears	5	tell + lies	2
treat + patient	5	cure + disease	2
cure + illness	5	have + doubt	2
make + decision	5	give + attention	2
ask + question	5	loose + hope	2
give + honour	4	put + trust	2
learn + lesson	4	Total	254

The table contains 41 different Congruent Verb Noun Collocational Structures which were used 254 times.

Table 7.13 Congruent Verb Noun Collocations Produced by B1 Learners

Congruent Verb Noun Collocations Produced by B1 Learners			
Collocations	Frequency	Collocations	Frequency
go to + church	13	write + letter	3
cut + cake	8	open + door	3
have + accident	5	play + music	3
pass + examination	5	have + opportunity	3
go to + bed	4	pour + water	2
collect + result	3	comb + hair	2
read + story	3	use + opportunity	2
tell + story	3	**Total**	**62**

There are 15 different Congruent Verb Noun Collocational Structures used 62 times.

Table 7.14 Congruent Verb Noun Collocations Produced by A2 Leaners

Congruent Verb Noun Collocations Produced by A2 Learners			
Collocations	Frequency	Collocations	Frequency
spend + holiday	36	pay + money	4
wash + plate	27	read + story	4
play + ball	20	watch + television	4
fetch + water	11	pass + examination	4
ask + question	10	ride + bicycle	3
go to + bed	10	meet + friend	3
brush + teeth	8	write + examination	2

tell + story	7	answer + question	2
have + experience	5		
go + lesson	4	**Total**	**164**

There are 18 different congruent verb noun collocational structures collectively used 164 times in the corpus.

7.4 Production of Incongruent and Congruent Adjective Noun Collocations across Proficiency Levels

All the four proficiency groups produced fewer adjective noun collocations in comparison to the verb noun collocations they produced. Unlike the verb noun collocations category above, all the four groups produced more congruent adjective noun collocations than incongruent ones. See table 7. 15 for the overall statistics of the adjective noun collocations and tables 7. 16, 7.17 and 7.18 for the details of the incongruent adjective noun collocations produced by the proficiency groups C1, B2 and B1. The A2 group produced no incongruent adjective noun collocations.

Table 7.15 Overall Adjective Noun Collocations

Overall Adjective Noun Collocations				
Proficiency Groups	Corpus Size	Adj. Noun Collocations	Incongruent Adj. Noun Collocations	Congruent Adj. Noun Collocations
NILECORP-C1	252,003	531	144	387
NILECORP-B2	130,559	222	23	199
NILECORP-B1	73,660	49	6	43
NILECORP-A2	66,996	27	0	27

Table 7.16 Incongruent Adjective Noun Collocations Produced by the C1 Group

Incongruent Adjective Noun Collocations			
Collocations	Frequency	Collocations	Frequency
secret + cult	28	last + night	4
bright + future	13	average + student	3
elder + brother	12	curious + mind	3
long + time	11	strong + influence	2
bad + disposition	11	pipe-borne + water	2
bouncing + baby	11	varying +degree	2
unwanted + pregnancy	10	tight + security	2
moral + education	6	deep + voice	2
nearest + future	5	bright + student	2
nonchalant + attitude	5	quick + reaction	2
social + miscreant	4		
peer + pressure	4	Total	144

Table 7.17 Incongruent Adjective Noun Collocations Produced by B2 Group

Incongruent Adjective Noun Collocations Produced by B2 Group			
Collocations	Frequency	Collocations	Frequency
fair + complexion	7	vital + role	3
medical + treatment	5	premarital + sex	2
sound + education	3		
wedded + wife	3	**Total**	**23**

There are six Incongruent Adjective Noun Collocational Structures used 23 times.

Table 7.18 Incongruent Adjective Noun Collocations Produced by B1 Group

Incongruent Adjective Noun Collocations Produced by B1 Group	
Collocations	Frequency
sweet + mother	4
bouncing + baby	2
Total	**6**

There are two Incongruent Adjective Noun Collocational Structures used two times.

Analysis of the incongruent adjective noun collocations produced by the four groups reveals that of all the adjective noun collocations produced by the C1 group, 27.1% of them were incongruent, for the B2 group, it is 10.3% incongruent, B1 group produced 12.2% while A2 group produced zero percent incongruent adjective noun collocations. Again, apart from the B1 group which has a result which is inconsistent with the rest, all the other three groups show a progression in the production of incongruent adjective noun collocations in tandem with proficiency increase. The B1 group only produced two different adjective noun collocational structures as can be seen in table 7. 18 above as opposed to the B2 group which produced six structures. So, they did not produce more than the B2 group.

Considering the congruent adjective noun collocations, the data again suggests the learners produced a lower percentage of congruent collocations as their proficiency increases. Starting with the least proficient, the A2 group produced only congruent adjective noun collocations. This suggests a one hundred percent reliance on L1 to produce adjective noun collocations. The most advanced group on the other hand produced 72.9% congruent adjective noun collocations. The B2 group which is the second most proficient group produced 87.8% while the B1 group produced 89.7%. Apart from the B1 group, all the other three proficiency groups consistently show that the learners produced more of incongruent adjective noun collocation and fewer of congruent ones as their proficiency increases. These findings are consistent with the earlier findings above. See tables 7.19, 7.20, 7.21 and 7.22 for more details on all the congruent adjective noun collocations produced by the proficiency groups

B2, B1 and A2. Meanwhile, I am representing the table showing the congruent adjective noun collocations produced by the C1 group, which I have presented earlier in chapter, for ease of analysis and comprehension.

Table 7. 19 Congruent Adjective Noun Collocations Produced by the C1 Group

Congruent Adjective Noun Collocations			
Collocations	Frequency	Collocations	Frequency
good + friend	56	second + wife	5
bad + behaviour	40	important + position	5
bad + character	30	bad + manner	5
peer + group	28	bad + reputation	5
best + friend	27	good + record	5
fellow + student	23	major + problem	5
bad + attitude	17	little + brother	4
bad + influence	15	good + character	4
armed + robbery	14	hard + work	4
bad + habit	10	true + friend	3
good + manner	9	several + times	3
final + examination	8	innocent + blood	3
best + student	7	age + group	3
class + mate	7	sweet + experience	2
good + care	6	secret + society	2
bad + language	6	evil + companion	2
close + friend	5	tight + friend	2
good + news	5	bad + news	2
main + purpose	5		
big + shame	5	**Total**	**387**

This table contains 38 different structures used 387 times in the corpus

Table 7.20 Congruent Adjective Noun Collocations Produced by the B2 Group

Congruent Adjective Noun Collocations Produced by B2 Learners			
Collocations	Frequency	Collocations	Frequency
best + friend	95	good + advice	4
true + friend	48	dark + complexion	4
good + girl	15	good + relationship	4
good + behaviour	8	armed robbery	3
strange + illness	5	many + advice	2
beautiful + girl	5	school + life	2
medical + attention	4	**Total**	**199**

There are 13 Congruent Adjective Noun Collocational Structures used 199 times.

Table 7.21 Congruent Adjective Noun Collocations Produced by the B1 Group

Congruent Adjective Noun Collocations Produced by B1 Learners	
Collocations	Frequency
younger + brother	15
senior + brother	8
younger + sister	5
new born + baby	4
elder + sister	4
best + friend	3
big + mother	2
painful + experience	2
Total	**43**

There are eight congruent adjective noun structures collectively used 43 times.

Table 7.22 Congruent Adjective Noun Collocations Produced by the A2 Group

Congruent Adjective Noun Collocations Produced by A2 Learners	
Collocations	Frequency
best + friend	12
younger + brother	6
younger + sister	6
elder + sister	3
Total	**27**

This table contains four congruent adjective noun collocational structures used 27 times.

Up to this point, the quantitative analysis has revealed a strong link between proficiency and the production of collocations. It is very interesting to see the reversal of relationship between the production of incongruent collocations and proficiency on the one hand and the production of congruent collocations and proficiency on the other hand. It seems plausible to conclude that L2 learners rely heavily on their L1 to produce L2 collocations, and this reliance thins out as they become more proficient. This results in their production of fewer incongruent collocations at the early stage of the acquisition process and increases as they become more proficient. Beyond this, in order to deepen our understanding of how proficiency reflects in the production of collocations, this analysis, from the next section, will focus on the linguistic complexity of the collocations produced across the four levels in terms of their collocations span and the structural and semantic properties of their constituents.

7.5 Linguistic Complexity of the Collocations Produced Across Proficiency Levels

This section will present the analysis the collocations produced by the four groups of learners in terms of the collocational span and the structural properties of their constituents. Only

verb noun collocations with three words or more will be considered in the analysis. All verb noun collocations that have articles or determiners like *my, his, her* etc. in between the collocate and the node will not be categorised as three-word collocations (e.g. '*compose my song', 'recite a poem'*). The aim of this section is to see how the production of structurally complex collocations reflects across the four proficiency groups. I will start with collocational span.

7.5.1 Collocational Span across Proficiency Levels

The data suggest more proficient learners produce collocations with a longer span than the least proficient learners. This is more pronounced between the C1 group and the A2 group and the gap reduces as the gap between the proficiency groups reduces. For instance, the difference between the collocations with long span produced by the C1 group and the B2 group is not as much as the difference between C1 and B1. There is an increase in the production of long span collocations in tandem with proficiency increase. But the A2 group which is the least proficient group produced longer span collocation than the B1 group which is one step above it. This was because of a repetitive use of certain structures by the A2 group.

Table 7.23 Statistics of Long Span Collocations Produced by the Learners

Long Span Collocations				
Proficiency Groups	Corpus Size	Total Verb Noun Collocations Produced	Long Span Collocations	Percentage
NILECORP-C1	252,003	793	194	24.46%
NILECORP-B2	130,559	377	74	19.62%
NILECORP-B1	73,660	164	9	5.48%
NILECORP-A2	66,996	234	31	13.24%

As can been seen in table 7.23 above, 24.46% of all the verb noun collocations produced by the C1 group were long collocational structures while 19.62% of the ones produced by the B2

group – the second most proficient group – are long span collocations. Three of the four proficiency groups consistently show an increase in the number of long span collocations as their proficiency increases. The only exception is the B1 group which produced fewer than the A2 group. Overall, it is plausible to infer that the acquisition of L2 collocations start with shorter strings of words [collocations] like *draw + conclusion* and as the learners become more proficient, they can then produce sometime like *draws an apparently illogical conclusion*. So, the production of long span collocations is indicative of proficiency as this data clearly suggest. The shorter strings of words like 'draw + conclusion' is more common in the input that learners are exposed to than complex longer strings of words like 'draws an apparently illogical conclusion'. As my earlier findings in chapter six suggest that learners are more likely to acquire the most frequent collocations before the least frequent ones, this group of learners seem to confirm that again by producing far more shorter strings of words which are more frequent. Besides, learners need to be more proficient to process the meaning of words before they can correctly use them in between the collocate and node to produce longer strings of collocations.

7.5.2 Structural Properties of Collocations Produced Across Proficiency Levels

Having established that the most proficient learners produce more long span collocations than the least proficient learners, the next phase in the analysis of the linguistic complexity of the verb noun collocations produced by the four learner groups is the structural properties of the collocations. The focus of this analysis is on the collocations that have collocations in their constituents (collocations within collocations). All the long span collocations produced by the groups were analysed and all the verb noun collocations that have collocations within their structures were isolated. The data suggests that least proficient learners have preference for less structurally complex collocations. This changes as their proficiency increases. The A2 and B1 groups produced three and two collocational structures respectively that have collocations in their constituents. Out of the 31 long span collocations produced by the A2 group, only three structures ('make *good friend'*, 'make some *good friend*' and 'told some *interesting stories'*) have collocations within their structure. And the structure 'good friend' was repeated twice. So, there are actually, only two collocational structures with collocation inside them. The B1, which is next least proficient group also produce only two structures

with collocation inside them. These structures are 'cut the *birthday cake*' and 'had a *tragic accident*'.

However, the two most advanced groups produced remarkably more structurally complex collocations than the two least proficient groups. The C1 group produced 34 collocations that have collocations inside their constituents. This is 17.5% of the 194 long span collocations they produced. See table 7.24 for more details on the complex collocations produced by the group.

Table 7.24 Collocation within Verb Noun Collocations in NILECORP-C1

Collocation within Verb Noun Collocations in NILECORP-C1			
write a *short story*	9	heed to a *good advice*	1
take *good care*	7	keeping *late night*	1
write this *short story*	4	make a *lot of money*	1
make *some money*	2	give *medical treatment*	1
make *quick money*	2	spend a *lot of money*	1
keeping *good record*	2	pay *less attention*	1
narrated the *whole story*	1		
learn a *good lesson*	1	Total	34

The B2 group produced 13 complex collocations. This 17.56% of the 74 long span collocations they produced. As we have seen in chapter five, native speakers use far more complex collocations – collocations with collocation as their constituents – in their written texts. What this data have suggested, as this phenomenon is investigated across proficiency levels is that as learners' proficiency increases, their use of complex collocations also increases. At the initial stage of acquisition, learners seldom use complex collocations. This will be discussed further in the discussion chapter. See table 7.25 for details of the complex collocations produced by B2 group.

Table 7.25 Collocation within Verb Noun Collocations in NILECORP-B2

Collocation within Verb Noun Collocations in NILECORP-B2			
give *medical treatment*	3	write their *final examination*	1
taken to a *general hospital*	1	write the *promotional exam*	1
give you a *good advice*	1	make any *real difference*	1
give her some *piece of advice*	1	make the *right decision*	1
give her a *word of advice*	1	make a good and *wise decision*	1
save *some money*	1	**Total**	**13**

7. 6 Semantic Properties of Collocations Produced Across Proficiency Levels

The collocations produced by the four proficiency groups were analysed for their semantic properties. The semantic properties here refer to the use of collocations along a continuum of decreasing or increasing semantic transparency and/or opacity. The analysis also includes collocations with delexical verbs such as take, make, have, etc. These verbs establish their meaning from the word (node) they are combined with. In this way, these verbs take on additional meaning and are therefore, semantically burdensome. It is agreed in the literature that for L2 learners, these verbs present difficulties when it comes to collocations (McCarthy, 2014). The aim of this section is to deepen our understanding of how L2 learners use collocations with modified meanings – with additional nuances and associations – across proficiency levels. All collocations with modified meanings were isolated and analysed.

There are 92 instances of verb noun collocations and 26 instances of adjective noun collocations with modified meanings in all the collocations extracted from NILECORP-C1 group which is the most proficient group. In total, they produced 118 collocations with varying degree of idiomaticity. They produced far more of these collocations than the other three groups. See table 7.26 for more details of the verb noun collocations in this data set.

Table 7.26 Verb Noun Collocations with additional meanings in NILECORP-C1

V + N Collocations with additional meanings in NILECORP-C1			
Collocations	Frequency	Collocations	Frequency
take + care	39	keeping + late night	2
take + bath	10	draw + attention	2
tarnish + image	9	tarnish + reputation	2
handle + equipment	7	turn down + offer	2
fall in + love	7	keep + record	2
pay + attention	5	shed + blood	2
damage + image	3	**Total**	**92**

As can be seen in the table above, such expressions as: 'tarnish + image' and 'tarnish + reputation' where 'tarnish' which literarily means 'lose or cause to lose lustre or dullness of colour or lost of brightness' is being used figuratively here in relation to the beliefs or opinions that are generally held about someone (reputation and/or image). In a similar way, the learners used the structures: 'fall in + love', 'turn down + offer' and 'shed + blood' – all these expressions have additional nuances and associations. All these six expressions, though not completely opaque, but could be considered to be towards the upper end of the continuum of semantic opacity. The structures 'fall in love', 'turn down offer' and 'shed blood' seem to be more semantically opaque. Besides, apart from 'shed + blood' which is congruent (has Yoruba equivalence), all the other structures are incongruent. Yet these groups of learners were able to use them correctly to convey figurative meanings.

Four different collocational structures have delexical verbs. All the verbs would mean something else if isolated from the nodes but by using them with the nodes, they have taking on additional meanings. These four structures are: 'keep + record', 'take + care', take + bath' and 'keeping + late night'. Consider 'keeping late night' which means habitually staying out till late in the night in Nigerian English, while 'late night' is transparent, the verb 'keeping' in this combination has had its meaning completely altered. In a similar way, all the other

lexical verbs in ‘keep + record’, ‘take + care’, take + bath’ have had their meaning modified. For instance, bathing is something to be done not to be taken. To ‘take bath’ means to ‘do’ the act of bathing.

Other verbs in the structures like ‘handle’ in handle equipment, ‘pay’ in pay attention, and ‘draw’ in draw attention have also taken on additional nuances. ‘Handle’ in combination with ‘equipment’ means to use (equipment). For L2 learners, the verb ‘pay’ will, in most cases at the initial stage of acquisition, be associated with money. Then as they become more proficient, they may be able to associate more meaning to the verb ‘pay’. For instance, they may know that ‘pay the price’ could mean more than payment of money. It could mean face the consequence of the bad things someone has done. What the data has clearly shown is the learners’ ability to use collocations to express figurative meaning. However, in comparison with native speakers’ use of collocations to convey figurative meaning as shown in Chapter Five, the degree of idiomaticity of the verb noun collocations produced by this learner group is not at the extreme end of the continuum of opacity. Notwithstanding, they have demonstrated appreciable mastery of the use of figurative collocations

Their use of collocations with additional meanings is not limited to verb noun collocations. Out of the 531 instances of adjective noun collocations in the NILECORP-C1, 26 of them have additional meanings (see table 7.27 below). The collocations have varying degree of idiomaticity. For instance, ‘innocent blood’ which means an innocent person seems more semantically opaque than the other structures in table 7.27, though also semantically opaque but with lesser degree of opacity. The adjective ‘bright’ in ‘bright student’ and ‘bright future’ has nothing to do with brightness of colours but in these combinations, it has put on added meaning. ‘Bright student’ means a student who is intelligent and quick to learn while ‘bright future’ may mean a promising future.

Table 7.27 Adjective Noun Collocations with additional meaning in NILECORP-C1

Adj. + N Collocations with additional meanings in NILECORP-C1			
Collocations	Frequency	Collocations	Frequency
bright + future	13	strong + influence	2
innocent + blood	3	deep + voice	2
bright + student	2	tight + security	2
sweet + experience	2	**Total**	**26**

The data suggests this group of learners are able to produce a substantial number of semantically burdensome collocations. If the collocations produced by this group are considered within a continuum of semantic transparency and opacity, they would be situated somewhere toward to upper end of semantic opacity. I will now consider the second most proficient group.

The B2 group produced only one adjective noun collocation which could be considered as having additional meaning. They produced the structure 'sound + education' three times in the learner corpus. 'Sound education' in Nigerian English means high quality education. The adjective 'sound' in this combination has had its meaning modified. We can see a sharp difference in the number of adjective noun collocations produced by the C1 group and the B2 group – 1 versus 26.

Out of the 377 verb noun collocations produced by the B2 group, only 27 have additional meanings (see table 7.28 below). There is some overlap in the collocations produced by C1 and B2 groups. Expressions like 'take care', 'take bath' and 'pay attention' are in both sub-corpora. Since I have analysed these overlapping collocational structures earlier, I will only focus on new structures in the analysis for this group. An examination of the collocations with additional nuances and associations in this data sub-set shows they are more semantically transparent compared to the ones produced by the most proficient learner group. The correctly used structures with delexical verb such as 'make + friend' and 'make

difference' which could be problematic for learners. The verb 'make' in both instances above has taken on additional meaning. Also, the 'sit' in 'sit + examination' seems to be a metaphor for writing an examination and so have acquired new meaning. The most semantically opaque structure is probably 'carry out + operation' which means the act of operating on someone in Nigerian English. The phrasal verb 'carry out' in that collocation conveys a meaning that is far removed from the literal meaning of those words. This data clearly suggests two things. One is that the numbers of collocational structures with additional nuances and associations reduces remarkably from C1 to B2 and the second is that the degree of idiomaticity of the expressions is quite low compared to the most proficient group. If I were to put such collocations produced by the B2 group in a continuum of semantic transparency and opacity, I would situate them somewhere toward to lower end of semantic opacity.

Table 7. 28 Verb Noun Collocations with additional meaning in NILECORP-B2

V +N Collocations with additional Meanings in NILECORP- B2			
Collocations	Frequency	Collocations	Frequency
make + friend	8	put + trust	2
take + bath	6	sit + examination	2
take care + patient	3	pay + attention	2
make + difference	2		
carry out + operation	2	**Total**	**27**

The outputs for the two least proficient groups are quite similar. Both of them produced no adjective noun collocations with modified meanings. There are 56 instances of these types of verb noun collocations in corpus produced by the B1 group while there are 52 instances of such collocation in corpus produced by the A2 group (see tables 7.29 and 7.30 for more details). Both groups produced four different verb noun collocational structures each. If we consider this in proportion to the size of the sub-corpora, then the A2 group produced more. But the difference is negligible. Both groups did not produce collocations that could be

regarded semantically opaque. However, they were able to use correctly collocations with delexical verbs. Again, there are some verb noun collocations in the two data sets that are also produced by the previous two groups. I will, therefore, analyse only the structures which I have not analysed in the earlier data sets. The structures I will analyse are: 'give + birth', 'take + picture', 'have + breakfast', and 'take + notice'. Just like the other delexical verbs, 'give' when combined with 'birth' have acquired additional meaning. It is the same with the verbs 'take' in 'take + picture', 'take' in 'take + notice' and 'have' in 'have + breakfast'. The verbs have acquired additional meaning. The key here is that all these structures have had additional semantic burden for the learner to process and produce.

Table 7.29 Verb Noun Collocations with additional meaning in NILECORP-B1

V + N Collocations with additional meaning NILECORP- B1			
Collocations	Frequency	Collocations	Frequency
give + birth	43	take + picture	2
take + bath	11		
have + breakfast	3	**Total**	56

Table 7.30 Verb Noun Collocations with additional meaning in NILECORP-A2

V + N Collocations with additional meanings in NILECORP-A2			
Collocations	Frequency	Collocations	Frequency
take + bath	34	take + notice	5
give + birth	10	have + bath	3
		Total	**52**

The data suggests that the use of collocations with figurative meanings is almost non-existent in the text produced by the least proficient groups – B1 and A2. However, as their proficiency increases, their use of collocations with additional nuances and associations increases. Two main themes have emerged from this result. The first one is the role of the age of the learners in the production of collocations with figurative meaning. And the second is whether the acquisition of collocations with figurative meanings mirrors the acquisition of L1 figurative language. The learners whose texts formed the NILECORP-A2 and NILECORP-B1 are between the ages of 13 and 14 while the learners whose texts formed the NILECORP-B2 and NILECORP-C1 are between 15 and 16. As this data suggests, 15 and 16 year olds have appreciable productive knowledge of collocations with figurative meanings while the 13 and 14 year olds clearly demonstrated deficiency of the productive knowledge of these types of collocations. Both in quantity and quality, the later groups stand out – they produced collocations that could be situated toward to upper end of the semantic transparency/opacity continuum while the former groups are at the bottom of the continuum. All these seem to suggest that at the initial stage of acquisition, learners first learn the literal meaning of collocational combinations and then then figurative meanings. But there is a caveat to this. Will the result be the same if we study adult L2 learners who are at the same proficiency levels? Learners who are by the virtue of their ages have a vast knowledge of figurative expressions in their L1 might transfer some of their L1 knowledge to produce L2 collocations with figurative meanings. In the bigger discussion chapter, I will explore the effect of age on the production of figurative expression to try to explain these findings.

7.7 Discussion

The inquiry into the relationship between language proficiency and the production of collocations has revealed many things. This discussion section aims to discuss these findings within the immediate literature. The discussion here previews a deeper discussion later in the discussion chapter within the wider literature on second language acquisition with more focus on L2 collocations acquisition. These findings corroborate earlier findings that L2 lexical competence and L2 collocational competence develop as proficiency of L2 learners increase (Nizonkiza, 2011, 2015; Zareva, Schwanenflugel & Nikolova, 2005). This study suggests a

strong link between language proficiency and the production of collocations in three out of the four proficiency groups. The fact that not all the four proficiency groups show a consistent positive relationship between proficiency and production of collocations suggest there is more to know about the relationship between proficiency and production of collocations. It could be that the fine-grained categories of proficiencies do not map onto the progress of learners in relation to collocations. But there seems to be a convincing progression in the production of collocations as learners' proficiency increases in this study. This is consistent with various findings in the literature. But this study is wider and deeper in scope than many previous studies. This study did not just look at the collocations produced in quantitative terms, but the linguistic quality of the collocations produced across four proficiency levels within the context of Nigerian English with its distinctive characteristics (and even its own collocations which may not be in other prestigious varieties such as the British English).

The wider scope of this study means a discovery of wider range of findings that seem, to the best of my knowledge, not to be in the existing literature. One of such findings is that while the production of incongruent collocations increases as proficiency increases, the production of congruent collocations decreases as proficiency increases. It is well documented in the literature that incongruent collocations are problematic for learners (Yamashita & Jiang, 2010; Peters, 2016). It is logical that as learners' language proficiency increases, they seem have more language knowledge to alleviate the problem of producing incongruent collocations. This may explain why their incongruent collocational output increases as their proficiency increases. What seems to be a new discovery is that the production of congruent collocations decreases as their proficiency increases. As I have said earlier, this may mean that as L2 learners become more proficient, they rely less on their L1 to produce L2 structures. Hence, their production of collocations which have no L1 equivalent increases in tandem with proficiency increase while their production of collocations which have L1 equivalent decreases as their proficiency increases. The less proficient L2 learners relying heavily on their L1 to produce L2 collocations would naturally be able to produce collocations that are congruent with their L1. This means producing more congruent collocations at the initial stage of acquisition but as their proficiency increases with more L2 lexical items in their linguistic repertoire, they can rely less on their L1 to produce collocations and thereby increasing their production of incongruent collocations.

Most of the studies reported in the literature on the relationship between collocations and language proficiency limit their comparison to the frequency and correctness of collocations produced across proficiency levels (Laufer and Waldman, 2011; Ebrahimi-Bazzaz et al, 2014; Talakoob & Koosha, 2017). This study, however, widens the scope to include length of the string of words forming the collocations (collocational span), and the structural and the semantic properties of the collocations. The findings suggest at the initial stage of acquisition, L2 learners produce more of two-word collocations. As their proficiency increases, their production of long span collocations increases. This seems to explain why the more proficient learners in this study produced collocations with longer span than the least proficient learners. This will be explored further in the wider discussion chapter.

Another theme that emerged in this chapter is that the production of structurally complex collocations is indicative of language proficiency. As the data clearly show, the two most advanced groups produced remarkably more structurally complex collocations than the two least proficient groups. The structural properties of collocations have been neglected in L2 collocations research apart from Bartsch (2004) who published a volume on the functional and structural properties of collocations. Her book – a corpus study of lexical and pragmatic constraints on lexical co-occurrence – however, was not a study of the structure of collocations in relation to proficiency. This thesis attempts to investigate how L2 learners at various proficiency levels navigate through the constraints on lexical co-occurrence to produce structurally complex collocations. To the best of my knowledge, there is no literature on this aspect of collocational acquisition. In my comparative analysis of complex collocations produced by native speakers and the most proficient of the four learner groups (NILECORP-C1) in this study in chapter five, the written text of native speakers contains a substantial amount of collocations that have collocations within their structure. The number of similar structures in the NILECORP-C1 is quite few in comparison to the native speakers. Comparing the production of such structures across proficiency levels reveals that at the initial stage of acquisition, learners produce less structurally complex collocational structures. As proficiency increases, they produce more complex collocational structures. Even then it might be difficult for L2 learners to produce as many complex collocational structures in their written texts as native speakers would do.

Another theme which is closely related to the structural properties of collocations which also emerged from this chapter is the semantic properties of collocations – the production of collocations with additional nuances and associations. A large body of literature exists on L2 collocational processing (Siyanova & Schmitt, 2008; Wolter & Yamashita, 2013; Yamashita & Jiang, 2010) with their focus, essentially on the effects of congruency and frequency of input apart from Gyllstad and Wolter (2015) who took semantic criteria into account. Their findings suggest "that semantic transparency affects processing of word combinations, both for NSs and NNSs; more specifically, when defined along the lines of the phraseological tradition, collocations were processed slower than free combinations" (ibid: 317). This means semantic transparency or opacity plays important role in the acquisition of collocations. The findings in this study suggest that the use of collocations with additional nuances and associations increases as proficiency increases. This seems to mean that the processing cost for such colocations is more pronounced at the initial stage of acquisition. But there could be another twist to these findings. As I have said earlier, the lack of use of collocations with additional meanings by the least proficient groups might not necessary be a function of their language proficiency, it might be because they are young. The literature on the production of L1 figurative expression could help to explain these. All these will be discussed further in the discussion chapter.

Chapter Eight

L2 Collocational Errors across Proficiency Levels

8.0 Introduction

In the previous chapters, the data suggested that Yoruba-speaking English learners at the proficiency level which is equivalent to C1 proficiency level of the Common European Framework of Reference for Languages can produce, in quantitative terms, as many collocations in their text as native speakers would do. But the collocations they produce differ substantially from the ones native speakers produce in terms of their structural and semantic properties. Further analysis has also suggested that the production of collocations increases in tandem with proficiency increase and that the frequency of collocations in the input positively impacts collocational processing and acquisition. The findings have also suggested that the most proficient learners produced more incongruent collocations than the least proficient learners and that the least proficient learners produced more of the collocations that are congruent with L1 while the most proficient learners produced fewer congruent collocations. However, what have not been investigated are the infelicitous collocational combinations the learners produced in their texts. These are collocational combinations that deviated from the acceptable norms in English. The issue of norms and standards in English language will be discussed extensively within the concept of World Englishes in chapter nine.

This chapter, therefore, inquiries into the collocational errors produced by the L2 learners. The focus is on the identification, classification and the analysis of all the erroneous verb noun and adjective noun collocations extracted from the Learner corpus. It addresses four broad questions: (1) What types of collocations are the most problematic for the Learners? (2) What is the nature and causes of the collocational errors in the Learner Corpus? (3) What are the similarities and differences in the error across proficiency levels? (4) What proportion of collocation errors are due to: [a] Inter-lingual factors and [b] Intra-lingual factors.

Deciding which collocations are erroneous in this study is not necessarily based on the notions of norms and standards of some of the prestigious varieties of English (British, American). But it is based on the sociolinguistic reality of language use in the Nigerian context. I mean Nigerian English – "new English, still in communion with its ancestral home but altered to suit its new African surroundings" (Achebe, 1975:62). It is important to emphasize this because the global spread of English; the emergence of New Englishes and the increasing use of English as a lingua franca for global communication mean it is necessary to reconsider how English language is described in terms of norms and standards. Achebe's assertion above lends credence to Seidlhofer's (2006: 1) argument that "speakers of English as a lingua franca (EFL) are beginning to conceptualize themselves not as exonormatively oriented learners of English but as legitimate speakers of a world language that is shaped by all its users". Though there are still a few voices in Nigeria who seem to continue to promote conservative British English norms (Ifecheobi, 2016), in reality, the English language usage in classroom discourse, in the media, in literary publications (across the three literary genres of prose, drama and poetry) is an amalgam of British English and Nigerian English with some intrusions from American English. However, outside the classroom, the norms we orient to in social interaction are almost entirely Nigerian English norms. It is well documented that the English language in Nigerian press reflects "lexical, structural and rhetorical features … that reveal a variety of English with a distinct Nigerian flavour" (Ehineni, 2014: 26). In another study of the English language in Nigerian press examining English idioms used in some Nigerian newspapers by Osoba (2014: 46) reveals that "the idioms have undergone modifications in the Nigerian press, breaking the rule of fixed collocation".

Still on English language usage, a linguistic stylistic analysis of educated Nigerian English conversation by Enyi (2015: 42) reveals that, "apart from the common core - features which it shares with the general conversational English, has some indexical markers which locate it in its socio-cultural and sociolinguistic context as English as a second language".

In view of the above, I henceforth use the term 'non-teacher norms collocations' to describe the 'erroneous' collocations instead of 'deviant' or 'non-native-like' which Nesselhauf (2005: 165) used. Using the term 'deviant' or 'non-native-like' may

suggest that the acceptability of the collocations is benchmarked on native English norms. But by using 'non-teacher norms', I have accounted for this sociolinguistic reality of language use in Nigeria because there are some expressions in Nigerian English that reflect a 'distinct Nigerian flavour' which Nigerian English teachers regard as acceptable but may not be acceptable in British English. These non-teacher norms collocations are generally not acceptable by Nigerian English teachers community – a community which I myself belong to.

This chapter is divided into five sections:

The first section presents the overall descriptive statistics of the data used for this study. This includes the overall number of non-teacher norms collocations extracted from each sub-corpus and the number of times each of such unacceptable structures is used in the corpus. The second section contains the parameters used for the classification and the analysis of all the erroneous verb noun collocations across the four proficiency levels respectively. The third section focuses on the non-teacher norms verb noun collocations in NILECORP-C1. This section is divided into four sub-sections. Each sub-section focuses on the unacceptable verb noun collocations produced by each learner group with the first sub-section further divided into two parts: one focusing on intralingual errors while the other focuses on interlingual errors. The fourth section, I present and analyse the data on the non-teacher norms adjective noun collocations produced by the four learner groups. This section is divided into four sub-sections. Each sub-section focuses on the unacceptable adjective noun collocations produced by each learner group.

While analysing the collocation errors, I will investigate the factors that may have induced the production of these collocational expressions. I will also consider the proportion of the non-teacher norms collocations in proportion to the overall collocations produced by the learners and try to understand what that means in terms of the collocational knowledge and development of the learners. Finally, in the last section, which is a discussion section, I will interpret and explain my findings and examine whether and how my research questions have been answered. The discussion will show how my findings relate to the immediate literature on L2 collocational errors.

8. 1 Overall Results

The descriptive statistics presented here describe the basic features of the data used to analyse the non-teacher norms collocations produced by the four learner groups. It provides simple summaries about the samples and measures used in this section. While all the collocations included in this study would have to occur, at least twice, in the learner corpus, all the instances of non-teacher norms collocations are included in the error analysis [including those that occur just once]. In NILECORP-C1, a total of 182 verb noun and adjective noun collocational combinations that deviated from what is acceptable in Nigerian English were extracted. If these non-teacher norms collocations are considered in relation to the well-formed collocations, these will be 13.7% (182 ÷ 1324 × 100 = 13.7). The next most proficient group, the NILECORP-B2, produced a total of 68 verb noun and adjective noun collocational combinations that deviated from what is acceptable in Nigerian English. Using the same formula, this represents 11.3%. In NILECORP-B1 and NILECORP-A2, 25 and 10 non-teacher norms collocations were extracted respectively. This is 10.7% for NILECORP-B1 and 3.8% for NILECORP-A2. See table 8.0 below for more details:

Table 8.0 Overall Statistics for Unacceptable Collocations

Proficiency Groups	Corpus Size	All Collocations	All Non-teacher norms Collocations	Percentage of errors
NILECORP-C1	252,003	1,324	182	13.7%
NILECORP-B2	130,559	599	68	11.3%
NILECORP-B1	73,660	233	25	10.7%
NILECORP-A2	66,996	263	10	3.8%

One striking thing this data reveals is that, contrary to my expectation, the more proficient the learners are, the more non-teacher norms collocations they produced. I had thought there would be more collocational combinations that deviated from Nigerian English in the least proficient learners' texts. This expectation was driven by the fact that the results of the comparative analysis of the production of collocations across proficiency levels in the

previous chapter suggested the more proficient the learners, the more collocations they produced. This data on the non-teacher norms collocational combinations seems to be suggesting that the same more proficient groups who produced the most well-formed collocations also produced the most non-teacher norms collocational combinations. The least proficient groups, on the other hand, who produced fewer well-formed collocations turn out to produce fewer non-teacher norms collocational combinations. I will discuss these findings further in the mini discussion section at the end of this chapter.

A further analysis reveals that 152 of the non-teacher norms collocational combinations are verb noun structures. If these non-teacher norms verb noun collocational combinations are considered in relation to the well-formed verb noun collocations, these will be 19.1% (152 ÷ 793 × 100 = 19.1). The next most proficient group, the NILECORP-B2, produced a total of 55 verb noun collocational combinations that did not conform to acceptable norms in Nigerian English. Using the same formula, this represents 14.5%. Twenty of non-teacher norms collocational combinations in NILECORP-B1 are verb noun structures while 7 of the ones in NILECORP-A2 are verb noun structure. This is 12.1% for NILECORP-B1 and 2.9% for NILECORP-A2. Again, the data suggests that as the production of verb noun collocations increases in tandem with proficiency increase, the production of non-teacher norms verb noun collocations also increase. See table 8.1 below for more details:

Table 8.1 Overall Statistics for Unacceptable Verb Noun Collocations

Proficiency Groups	Corpus Size	Verb Noun Collocations	Non-teacher norms V+N Collocations	Percentage of errors
NILECORP-C1	252,003	793	152	19.1%
NILECORP-B2	130,559	377	55	14.5%
NILECORP-B1	73,660	164	20	12.1%
NILECORP-A2	66,996	234	7	2.9%

The data for non-teacher norms adjective noun collocational combinations reveals something slightly different. All the four learner groups produced fewer infelicitous adjective noun combinations in comparison to the verb noun structures the produced. They also produced fewer well-formed adjective noun collocations in comparison to the verb noun collocations produced. This might not be a function of language deficiency but rather it could be that there are more verb structures in the corpus than adjectives.

Only 31 of the non-teacher norms collocational combinations in NILECORP-C1 are adjective noun collocational combination. If these non-teacher norms adjective noun collocational combinations are considered in relation to the well-form adjective noun collocations, this will be 5.8% (31 ÷ 531 × 100 = 5.8). The next most proficient group, the NILECORP-B2, produced a total of 13 adjective noun collocational combinations that are deviant from Nigerian English language norms. Using the same formula, this represents 5.8%. The two least proficient groups produced five and three non-teacher norms adjective noun collocational combinations respectively. If we consider this in proportion to the well-formed adjective nouns they produced, this will be 10.2% for NILECORP-B1 and 10.3% for NILECORP-A2. Their production of infelicitous adjective noun combinations did not seem to increase in tandem with proficiency increase. See table 8.2 below for more details.

Table 8.2 Overall Statistics for Unacceptable Adj. Noun Collocations

Proficiency Groups	Corpus Size	Adj. Noun Collocations	Non-teacher norms Collocations	Percentage of errors
NILECORP-C1	252,003	531	31	5.8%
NILECORP-B2	130,559	222	13	5.8%
NILECORP-B1	73,660	49	5	10.2%
NILECORP-A2	66,996	29	3	10.3%

The overall data seems to suggest that verb noun structures are more complex to produce considering the span and the structural complexity of their constituents. This could explain why these learner groups produced more non-teacher norms verb noun structures than adjective noun structures. The scale of the infelicitous collocations produced by the learners

is consistent with various findings in the literature. It is generally acknowledged that collocational deficiency is a pervasive phenomenon in second language learning (Biskup, 1992; Bahns, 1993; Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Durrant & Schmitt, 2009; Laufer & Waldman, 2010; Yamashita & Jiang, 2010; Boers, Lindstromberg & Eyckmans, 2014). The findings in this section will be explored further in the discussion section. In the meantime, the data on the classification and analysis of the non-teacher norms collocational combinations will be presented in the next section.

8.2 Classification and Analysis of Verb Noun Collocational Errors

This section focuses on the classification of the collocational errors and detailed analysis of the errors. The classification of the collocational error is based on the possible interpretation of the origin of the errors. The errors are classified into two broad categories namely: interlingual errors and intralingual errors. The errors classified as interlingual are caused by negative crosslinguistic influence while the ones categorised as intralingual are caused by negative transfer within the target language (Lim, 2007). This section is divided into four sub-sections focusing on the non-teacher norms verb noun structures in NILECORP-C1, NILECORP-B2, NILECORP-B1, and NILECORP-A2.

8.2.1 Non-Teacher Norms Verb Noun Collocational Structures in NILECORP-C1

The data on the unacceptable verb noun collocations produced by the learners are presented and analysed starting with the NILECORP-C1 learner sub-corpus. This group of learners produced 27 different unacceptable verb noun collocational structures. Interestingly, all the 28 structures except one are incongruent. This seems to be overwhelming evidence that incongruent collocations are problematic for these relatively advanced learners of English in a context where English is a second language. This is consistent with various findings in the literature that incongruent collocations are the most problematic for L2 learners (Laufer & Waldman, 2011; Walter & Gyllstad, 2011, 2013; Peters, 2016; Lee, 2016). The analysis of the well-formed verb noun collocations produced by this group of learners in chapter seven

shows that 43.3% of them are incongruent and 56.7% congruent. They obviously produced a substantial number of incongruent collocations but the data on collocational errors shows the difficulty they went through in producing incongruent collocations. They have produced 345 well-formed verb noun incongruent collocations and144 unacceptable verb noun collocational combinations – that is almost as many as half of the well-formed verb noun collocations. This suggests they have almost 50% possibility of producing unacceptable incongruent verb noun collocations. This highlights L2 learners' difficulty in producing incongruent collocations. This is not the case for the production of congruent verb noun collocations. The learners had produced 448 well-formed congruent verb noun collocations which is 56.7% of the well-formed verb noun collocations produced. Of all the congruent verb noun collocations they produced only one is unacceptable. This suggests that congruency have positive effect on collocation processing and acquisition. Besides, the only one unacceptable congruent verb noun collocation ('talk + story') produced by the learner lies in the borderline between congruent and incongruent collocation. This will be explained further while analysing the unacceptable collocations.

Meanwhile, the learners produced 27 different non-teacher norms verb noun collocational combinations which were collectively used 144 times in NILECORP-C1. See table 8.3 below for a list of all the unacceptable verb noun collocational combinations.

Table 8. 3 Non-teacher norm Verb Noun Collocations in NILECORP-C1

Non-teacher norm Verb Noun Collocations in NILECORP-C1			
Collocations	Frequency	Collocations	Frequency
join + bad company	46	join + bad friend	2
choose + friend*	18	abort + children	2
impact + knowledge	10	neglect+ advice	2
moving with + bad company	9	contact + disease	2
listen + instruction*	8	disobey + regulation*	2
get + accident	6	disobey + rules + regulations*	2

take + position*	5	condemn + image	1
talk + story	4	learn + habit	1
learn + work	4	exercise + view	1
conceive + children	3	put + prison	1
abort + baby*	3	abide + regulation	1
make + relations	3	attend + abortion	1
fight + riot	3	do + mistake	1
follow + gang	3	**Total**	**144**

[27 different non-teacher norm verb noun collocational structures used 144 times]

*these combinations are regarded as unacceptable because of the context in which they were used.

Of these 27 different unacceptable structures, 14 are categorised as interlingual errors while 13 are categorised as intralingual. The 14 structures in the interlingual errors category were produced 107 times in the learner corpus while the 13 structures in the intralingual category were produced 37 times. This means errors that result from negative L1 transfer constitute 74.3% of all the non-teacher norms verb noun collocations produced by this group of Yoruba-speaking learners of English. This finding is consistent with previous studies (Nesselhauf, 2003; Yamashita & Jiang, 2010) in the literature which attribute collocational errors mainly to negative crosslinguistic influence. This current data on the unacceptable verb noun collocations produced by the group indicates that intralingual factors account for 25.7%. Some of the errors classified as intralingual are phonological in the sense that they are caused by the learners' inability to discriminate the sounds of the co-occurring word(s) resulting in the substitution of collocating words with another word that sounds similar, e.g. *contract* and *contact*. Others seems to be caused by lexical deficiency in which the learners' limited knowledge of the vocabulary of English hindered complete and clear expression of idea. This results in the learners combining the co-occurring word with other words in a lexical set which would naturally not be used together. I will now analyse each of the unacceptable verb noun structures produced by the learners. At the end of the error analysis, we should be able to find out what proportion of the errors is interlingual and what proportion is intralingual as the data are analysed across the four proficiency levels and to see how proficiency might affect this.

8.2.1.1 Interlingual Verb Noun Collocational Errors in NILECORP-C1

The following verb noun structures are the non-teacher norms collocations. They are all heavily influenced by Yoruba language – the learners' L1. All the expressions seem to be a direct translation from Yoruba to English. In the absence of direct Yoruba equivalent of this expression, the learners seem to resort to creating the structures relying on their knowledge of L1 structure but the resultant combinations, though intelligible to Nigerian English speaker, they would be picked up by English language teachers as incorrect in the classroom.

join + bad company	follow + gang	learn + work
choose + friend	join + bad friend	fight + riot
moving with + bad company	disobey + regulation	learn + habit
listen + instruction	disobey + rules + regulations	put + prison
take + position		do + mistake

The common expression in Nigeria is 'keep + company', 'keep + gang', 'keep + bad friend'. These expressions, as I have said earlier are incongruent. While the group of learners produced 'keep + gang' 10 times, selecting the acceptable verb 'keep', they, however, produced four structures above in which they could have used the verb 'keep'. The meaning of 'keep bad company', 'keep gang' or 'keep bad friend' in Nigerian English is literally to start going out with bad people. If we were to interpret that in Yoruba, it would mean to 'join' (add yourself), 'move' (to start going about with), or to 'follow' (to follow someone's lead). So, the expressions: 'join bad company', 'moving with bad company', 'follow gang' and 'join bad friend' have their origin in Yoruba which is directly transferred to English. Besides, the fact that the learners produced 'join bad company' 46 times shows the extent to which L2 learners rely on their L1 to produce incongruent collocations.

Another striking instance of L1 interference is the production of 'choose + friend' which was produced 18 times. Choosing friends could be a correct expression in English, but these learners used the verb 'choose' in contexts where it was more appropriate to use the verb 'make' as in 'make friend'. There is no Yoruba equivalent of 'making friend', the act of

making friends in Yoruba is to 'have' (possessing) friends or 'choose' (select and acquire) friends. This explains why they select 'choose' friend instead of make friend. However, this group of learners produced 'make + friend' structures correctly 30 times. That means there were 48 instances in the corpus where the appropriate collocate would be 'make' and the learners got it right 30 times but got it wrong 18 times. Frequency data from the Nigerian component of GloWbE indicates that the collocation 'make + friend' is a frequent expression in Nigerian English. But the fact that the learners got this collocation wrong 18 times despite it being a frequent expression highlights the difficulty learners have producing incongruent collocations.

The learners also used 'disobey' (rules and) regulations four times. This stems from their direct translation of the Yoruba equivalent of 'break + (rules and) regulations'. To break the law or rule and regulations in Yoruba language means to 'disobey' – failing to comply with law and rules and regulation. Another example of negative L1 transfer are the non-teacher norms structures: 'learn + work' and 'learn + habit'. In Yoruba language, if someone is learning a trade, it is 'o n ko ise'. 'ko' means learn while 'ise' means work. However, the 'trade' in the structure 'learn + trade' means a job that needs special skills, especially the one that involves using your hand. In a similar way, the Yoruba language describes the formation of habit as something to be learned like learning a trade hence the learners produced 'learn + habit'.

Further analysis shows that when the learners used 'listen + instruction', the appropriate combination is 'follow + instruction'. Semantically, when someone says in Yoruba 'listen' to my instruction, they mean 'follow' my instruction. This is another case of L1 transfer negatively affecting the resultant combination. Some of these deviations are benign and may not result in communication breakdown even with an audience that is not familiar to the Nigerian communicative context. One non-teacher norm collocation that may be unintelligible to non- Yoruba speakers is 'take + position'. The position in this context means something like first position, second position, third position, etc. In the Nigerian educational systems, students are graded as having first position, second position, etc. This position in Yoruba language, is 'ipo' and to be in 1st, 2nd or 3rd position for instance, is described in Yoruba as 'gbe ipo ikini, ikeji abi iketa' which literally means 'to carry or take

1st, 2nd, or 3rd position'. So, the production of 'take + position' which was produced five times in the learner corpus is induced by the learners' L1.

The other three unacceptable combinations are also heavily influenced by the Yoruba language. The combination 'put + prison' for instance emanates from the Yoruba equivalent of 'sentenced to prison'. If someone is sentenced to prison, the Yoruba will say 'ju si inu ewon' or 'so si inu ewon'. The Yoruba verb 'ju' and 'so' mean to throw. The expression: 'ju si inu ewon' or 'so si inu ewon' which literally means to 'throw into prison'. Another Yoruba verb that can be used instead of those two verbs in relation to being sentenced to prison is the verb: 'fi si' as in 'fi si inu ewon' which means 'put in prison'. Similarly, in Yoruba, if some people are rioting, the verb to describe it means more of fighting. This explains why the learners combine 'fight + riot'. Finally, in the analysis of the interlingual errors, I will analyse the combination 'do + mistake'. The equivalent of the verb 'make' and 'do' in Yoruba is 'se' which fits in more in the context where we will use the verb 'do' in English. This might have influenced the learners' choice of 'do + mistake'. This data has shown the extent to which Yoruba language influences their production of incongruent verb noun collocations. I will now analyse the intralingual errors.

8.2.1.2 Intralingual Verb Noun Collocational Errors in NILECORP-C1

The non-teacher norms collocational structures below will be analysed in this sub-section. These are collocational errors which I refer to as intralingual emanate from within the L2 English.

condemn + image	neglect+ advice	exercise + view
conceive + children	contact + disease	abide + regulation
abort + baby	impact + knowledge	attend + abortion
talk + story	make + relations	get + accident
abort + children		

Two out of the thirteen unacceptable combinations in the category can be attributed to the learners' inability to discriminate the sounds of the co-occurring word(s) resulting in the substitution of the collocating words with another word that sounds similar. The learners

have mistaken ‘contract disease’ for ‘contact disease’ and ‘impart knowledge’ for ‘impact knowledge’. Their apparent inability to discriminate the sounds of these words (contract/contact and impart/impact) have resulted in the selection of the wrong collocate. This type of error may not lead to communication breakdown particularly in oral conversation. It may not even be noticeable. But that is not the case for errors like: ‘conceive + children’, ‘abort + baby’ and ‘abort + children’. These combinations seem to be caused by lexical deficiency in which the learners’ limited knowledge of the vocabulary of English hindered complete and clear expression of idea. The learners seem to combine words that belong to what looks like a lexical set – children, baby, conception, abortion. But while it is acceptable to say: abort pregnancy, it is infelicitous to say: ‘abort baby’ or ‘abort children’ as children/baby and pregnancy are not the same. Pregnancy can result in children/baby but while you can abort pregnancy, you cannot abort children/baby. The fact that the learners used related words suggest they have the receptive knowledge of the correct collocation: ‘abort pregnancy’ but lack the productive knowledge. This tends to confirm various studies that L2 learners’ productive knowledge of collocations lags behind their receptive knowledge (Talakoob & Koosha, 2017). Besides, these erroneous collocational expressions reveal the complexity involved in the production of incongruent collocations.

Meanwhile, it seems the learners wanted to produce the partially figurative collocation: ‘destroy + image’ but instead produced: ‘condemn + image’. However, they produced ‘destroy + image’ four times in the corpus which is acceptable. By producing ‘condemn + image’ suggests they have the receptive knowledge of the collocation but have difficulty producing it. This could be because it is incongruent and not entirely semantically transparent. They seem to have thought ‘condemn’ could substitute ‘destroy’ in this collocation. In the same vein, they appear to have substituted ‘build’ with ‘make’ in ‘build + relations’ and have produced ‘make + relations’ which is infelicitous in the Nigerian context. The same thing seems to have happened in the production of ‘talk + story’ where the learners appeared to have used ‘talk’ as a synonym of ‘tell’ thereby producing ‘talk + story’ instead of ‘tell + story’. The Yoruba equivalent verb for ‘tell’ and ‘talk’ is ‘so’ while story is ‘itan’. To tell a story would be ‘so itan’. While you can use the Yoruba verb ‘so’ in both the context where English will use ‘tell’ and ‘talk’, you cannot use ‘so itan’ as ‘talk + story’. This collocation is congruent and should not be problematic to produce but it seems the learners are confused by the verb ‘so’ meaning both ‘tell’ and ‘talk’. It could also be that the learners

used ‘talk’ as the synonym of ‘tell’ in this collocation. There is also the case of the polysemous nature of collocation making it difficult for the learners to produce the correct collocation (Pavlenko, 2009; Macis & Schmitt, 2016). I use polysemy from the perspective of combinatorial properties of lexical units to mean the capacity of a co-occurring word to have more than one meaning. I will discuss the polysemous nature of collocations in the discussion section at the end of this chapter. So, the error I have discussed above is both intralingual and interlingual. The following unacceptable structures: ‘neglect advice’, ‘exercise + view’, ‘abide + regulations’, ‘attend + abortion’ and ‘get + accident’ seem to have resulted from the learners being adventurous with the use of the English language but in the process producing infelicitous combinations. All these errors seem to emanate from within the English language.

8.2.2 Non-Teacher Norms Verb Noun Collocational Structures in NILECORP-B2

This group of learners, the second most proficient group, produced 16 different unacceptable verb noun collocational structures which were collectively used 49 times in NILECORP-B2. All the unacceptable structures are incongruent. This again suggests that incongruent collocations are problematic for L2 learners. See all the unacceptable collocational structures in table 8.4 below:

Table 8.4 Non-Teacher Norm VN Collocations in NILECORP-B2

Non-Teacher Norm VN Collocations in NILECORP-B2			
Collocations	Frequency	Collocations	Frequency
impact + knowledge	21	create + relationship	1
describe + drug	7	Improve + rate	1
give + lesson	5	rendering + attention	1
disagree + motion	2	do + advice	1
contact + disease	2	make use + advice	1
score + dream	2	get + accident	1

gather + knowledge	1	make + sex	1
inculcate + knowledge	1		
hold + attention	1	**Total**	**49**

16 different collocational structures used 49 times.

All the collocational errors in this sub-section are intralingual apart from two structures which seem to be attributable to the learners' L1 interference. Three out of the 16 structures are also among the errors in NILECORP-C1. These structures are: 'impact + knowledge', 'contact + disease', and 'get + accident'. As stated earlier, these errors are induced by the learners' inability to discriminate the sounds of impact/impart and contract/contact while the learners seem to mix-up the verb 'get' and 'have' in producing 'get + accident' instead of 'have + accident'. These learners also produced 'describe + drug' seven times in the corpus. This is another case of the inability to discriminate the sound of prescribe and describe. All these are cases of phonological errors. There are, however, eight instances where they produced the correct structure: 'prescribe + drug'.

A study by Farghal and Obiedat (1995:315) reveals that, L2 learners "heavily resort to strategies of lexical simplification like synonymy, paraphrasing, avoidance and transfer" because of their collocational deficiencies. This is what seems to happen when this group of learners produced: 'gather + knowledge' and 'inculcate + knowledge'. While it is natural to say 'acquire + knowledge' or 'acquire + wealth' in Nigerian English, saying 'gather + knowledge' or 'inculcate + knowledge' is not. The learners seem to have resorted to the strategy of using synonym to overcome the hurdle of producing this collocation. They seem to have mistaken 'gather' as a synonym of 'acquire' and the resultant combination is unacceptable. The production of 'inculcate + knowledge' seems to be the case of overgeneralisation. The verb 'inculcate' frequently co-occurs with various nouns like: 'inculcate + values', 'inculcate + discipline', 'inculcate + habit', 'inculcate + ideas', etc. in Nigerian English according to the frequency data from the Nigerian component of the Corpus of Global Web-Based English (GloWbE). The learners might have been exposed to the use of *inculcate* co-occurring with these nouns. What they did not seem to realize is that *inculcate* cannot naturally collocate with certain nouns even if their meaning is closely related to any of the above nouns e.g. 'knowledge' and 'ideas'.

Using synonyms seems to be the learners' most preferred way of getting around collocational difficulties. They seem to have used that strategy in producing the following combinations: 'disagree + motion', 'create + relationship' and 'get + accident'. They seem to have used 'disagree' instead of 'oppose' (oppose + motion), 'create' instead of 'build' (build + relationship) and 'get' instead of 'have' (have + accident). Using a verb that is seemingly synonymous in the above means falling foul of restrictions on the co-occurrence of words. While the expressions are intelligible, they are essentially, deviant sequences from the norms of Nigerian English. It seems the learners also resort to using synonyms in the production of 'make + sex'. The common acceptable collocation is 'have + sex' but the learners' use of the verb 'make' may stem from the concept of 'to do' as in 'doing sexual act' and 'making sexual act'. The combination: 'give + lesson' may be a result of the learners using 'give' in place 'provide' as in 'provide + lesson' (provide tutorial). All the errors analysed so far are intralingual.

However, there are two expressions in this dataset that seem to be interlingual. The expressions: 'do + advice' and 'make use + advice' seem to have their origin in Yoruba language. The most acceptable way of saying what the learners wanted to say would have been 'follow + advice'. To say 'follow my advice' could be expressed in three common ways in Yoruba language. One could say: 'se bi mo ti gba e ni imoran' (do as I have advised you). In this case, 'se' means to do and 'imoran' means advice. One can also say: 'mu imoran mi lo' which literally means 'make use of my advice' and the third common way of saying it is: 'te le imoran mi' (follow my advice). Of all the three common ways, only the third one is congruent with the acceptable English equivalent. The learners' production of 'do + advice' and 'make use + advice' must have had their origin from the other two common ways of 'saying + follow' advice in Yoruba.

While almost all the sources of the errors in this dataset could be identified and analysed, a few of the errors in this category are unexplainable. It is difficult to identify the source of the following errors: 'score + dream', 'hold + attention', 'improve + rate', and 'rendering + attention'. It could be that the learners considered dreams and goals as synonymous and therefore thought since it was acceptable to say 'score + goal' it should be acceptable to say 'score + dreams'. If that was the case, it is not natural to say 'score + goal' if that refers to

the object of someone's ambition or effort. But whatever the case, this data has revealed the extent of the complexity of processing collocations that have no L1 equivalent. This group of learners have resorted mainly to using synonymy as a strategy of overcoming the difficulty of producing incongruent verb noun collocations. This is consistent with various findings in the literature (Farghal & Obiedat, 1995; Shih, 2000; Davoudi & Behshad, 2015).

8.2.3 Non-Teacher Norms Verb Noun Collocational Structures in NILECORP-B1

This group of learners, the second least proficient group, produced nine non-teacher norms verb noun collocational structures. These structures were used 20 times in NILECORP-B1. All the collocational structures have no L1 equivalent. Meanwhile, as the data suggested in chapter seven, this is the only group that produced more incongruent verb noun collocations than the congruent ones. They produced 62.1% incongruent verb noun collocations. All the other groups had produced more congruent collocations than incongruent. Notwithstanding this achievement, the fact that all the unacceptable verb noun collocational structures they produced are incongruent suggests they have difficulty producing incongruent collocations. Out of the 20 instances of unacceptable collocations, 9 of them are interlingual while 11 are intralingual. See table 8.5 below for more details:

Table 8.5 Non-Teacher Norm VN Collocations in NILECORP-B1

Non-Teacher Norm VN Collocations in NILECORP-B1	
Collocations	Frequency
off + fire	7
do + holiday	6
do + cake	1
switch off + fire	1
vacate + holiday	1
climb + horse	1
drive + bicycle	1
wash + teeth	1
aboard + train	1
Total	**20**

Nine different non-teacher norms verb noun collocational structures used 20 times.

The most frequently used unacceptable structure is 'off + fire' which was used seven times in the learner corpus. This unacceptable combination seems to be attributable to the learners' lexical deficiency in which their limited knowledge of the vocabulary of English hindered complete and clear expression of idea. Combining *off* with *fire* to refer to the act of extinguishing fire suggests the learners had an idea of the correct collocation which is 'put out + fire'. But most likely, due to their lexical deficiency, they omitted 'put' and replaced 'out' with *'off'* which would have given the 'off' in their combination some meaning. It seems meaningless to say 'off + fire' without the 'put' and 'out' in this context. The other collocational combination in the learner corpus that relates to the extinguishing of fire is 'switch off + fire'. This is another case of collocational error apparently caused by the learners' lexical deficiency. While it is acceptable to say, for instance, 'switch off the light', it is not acceptable to say, 'switch off the fire'. The production of 'vacate + holiday' is another case of lexical deficiency. The learners might have established some connection

between vacation and holiday in their mental lexicon. This is probably why the learners wrote in the corpus that "I went to my sister when we *vacated on holiday*". Similarly, the learners are probably familiar with the collocation 'drive + car' and thought 'drive' can collocate with bicycle hence producing 'drive + bicycle'. It seems the learners' lexical deficiency is also responsible for the production 'aboard + train'. The collocation was used as follows: "my family woke up in the morning to *aboard a train*". They do not seem to understand the difference between board (as in board a train, which is the acceptable collocation) and aboard (as in he is aboard the train). All these errors seem to have their sources within the English language.

But this is not the case for the other errors. The structures: 'do + holiday', 'do + cake', 'climb + horse', and 'wash + teeth' are interlingual errors. They clearly have their root in Yoruba language. If you go on holiday, in Yoruba language, we would '… se isinmi'. The verb 'se' is the equivalent of the verb *to do* in English. This is the origin of the expression 'do + holiday' which these Yoruba-speaking learners of English produced six times in NILECORP-B1. This shows how much L2 learners rely on their L1 to produce incongruent collocations. Similarly, the verb 'bake' as in bake cake, has no equivalence in Yoruba. The verb to describe the act of baking cake in Yoruba is 'se' which is the same thing with the verb 'do' in English. So, while describing the act of baking cake in the learner corpus, they got it right twice and wrong once. They produced 'bake + cake' twice and 'do + cake' once which is a direct translation from Yoruba. The learner also produced 'ride + horse' correctly six times in the corpus. But there is one instance of a direct translation from Yoruba which resulted in the production of 'climb + horse'. In Yoruba, to ride a horse is 'gun esin' ('gun' means to climb while 'esin' is a horse). Finally, the verb 'brush' in brush teeth has no equivalence in Yoruba. Though the learners produced the correct collocation (brush + teeth) eight times in the corpus, there is still one instance of L1 interference. In Yoruba, we say, 'fo eyin' (wash teeth) which explains the production fo the combination: 'wash + teeth' instead of 'brush + teeth.

8.2.4 Non-Teacher Norms Verb Noun Collocational Structures in NILECORP-A2

The least proficient of the four learner groups produced only seven unacceptable verb noun collocational combinations. As is the trend in this data, the higher the proficiency, the more verb noun collocations they produced and the more unacceptable structures they produce as well. My expectation was that the more proficient they become, the fewer collocational errors they would make. This is quite interesting, and I will discuss it later in the discussion section at the end of this chapter. Meanwhile, all the collocational errors produced by this group of learners are incongruent. They produced four different acceptable verb noun collocational structures which were used all together seven times in the corpus. See table 8.6 below for more details:

Table 8. 6 Non-teacher Norm Collocations in NILECORP-A2

Non-teacher Norm Collocations in NILECORP-A2	
Collocations	Frequency
do + holiday	3
wash + television	2
talk + story	1
Started having + friends	1
Total	7

Four different non-teacher norms verb noun collocational structures used seven times.

The most frequently used of the unacceptable collocations is: 'do + holiday'. This structure was used across the two least proficient groups. The B1 group used it six times and this group (A2 group) used it three times. As have been said earlier in the analysis of the non-teacher norms verb noun collocational structures in NILECORP-B1, this error is a negative transfer from Yoruba language. The second error: 'wash + television' can be attributed to the learners' inability to discriminate the sounds of the co-occurring word(s) resulting in the

substitution of the collocating words with another word that sounds similar. The learners seem to have confused the sound of *watch* for *wash* which resulted in the production of 'wash + television' instead of 'watch + television'. This type of phonological factor has been the source of many errors in the learner corpus. Besides, this type of error could also be viewed as caused by learners' lexical deficiency. If they had had enough knowledge of the vocabulary of the English language, they should be able to differentiate the meaning of *wash* from *watch* and would not have used wash for television instead of watch television. I will address this phonological factor further when discussing L2 mental lexicon and its relationship with the production of clang associations – responses that have phonological resemblance to the stimulus words (Meara 1978, 1983; Namei, 2004) in the discussion chapter.

The third unacceptable structure – 'talk + story' – which was also used in NILECORP-C1 of is a negative transfer from the learners' L1. It may also be that the learners used 'talk' as a synonym of 'tell'. (refer to the section on Intralingual Verb Noun Collocational Errors in NILECORP-C1 for more details). The fourth structure in this dataset is not necessarily unacceptable expression but I have included it in this category because the learners seemed to be using paraphrasing to avoid producing the right collocations. They could have said: 'making friends' instead of 'started having friends. L2 learners have been found to avoid producing collocations by paraphrasing their way through (Farghal & Obiedat, 1995).

Finally, 56 different collocational structures have been analysed. These structures were used 220 times with very few of them repeated across the four proficiency groups. In essence, 220 instances of unacceptable collocations were analysed within the context of their usage. Out of these figures, 115 representing 52.2% of all the unacceptable verb noun collocations produced by the four groups of learners are L1-induced (interlingual) while 105 representing 47.8% are intralingual errors. All the L1-induced errors are a result of direct translation from Yoruba language. The intralingual errors are, however, cause by synonymy, paraphrasing, inability to discriminate sounds, and lexical deficiency. This means L1 negative transfer is the biggest source of errors in the production of L2 verb noun collocation. The analysis of the non-teacher norms verb noun collocations produced by the four learner groups also clearly shows incongruent collocations are problematic for learners. It further shows an

increase in the number of unacceptable collocations as the production of collocations increases.

8.3 Classification and Analysis of Adjective Noun Collocational Errors

The learners produced far fewer unacceptable adjective noun collocations in comparison to the volume of unacceptable verb noun collocations produced. In total, there are 48 instances of unacceptable adjective noun collocations in the four learner sub-corpora compared to 220 instances for verb noun collocations. Meanwhile, classifying the errors is not as straightforward as it is for the verb noun structures due to the difficulty in trying to identify the sources of the adjective noun collocational errors. The most workable strategy is to classify them based on the possible interpretation of the origin of the errors. Where there is obvious case of L1 negative transfer, they are regarded as interlingual errors and every other error even when the origin cannot be clearly identified, are regarded as intralingual errors. I will present the data and analyse them starting with the most proficient group.

8.3.1 Non-Teacher Norms Adjective Noun Collocational Structures in NILECORP-C1

This group of learners produced the highest number of non-teacher norms adjective noun collocations. They produced 20 unacceptable adjective noun structures which were collectively used 30 times in NILECORP-C1. See table 8.7 for more details.

Table 8.7 Non-teacher norm Adjective Noun Collocations in NILECORP-C1

Non-teacher norm Adjective Noun Collocations in NILECORP-C1			
Collocations	Frequency	Collocations	Frequency
parental + image	4	notorious + behaviour	1
junior + brother*	3	restful + mind	1
lunatic + attitude	3	uncomfortable + mind	1

toxic + friend	3	senior + brother*	1
nutritious + character	2	abysmal + attitude	1
vicious + friend	1	crucial + accident	1
negative + friend	1	exciting + mood	1
excretory + money	1	well-nurtured + manner	1
stubborn + character	1	unshakable + focus	1
paramount + behaviour	1	lunatic + acts	1
Many + money	1	**Total**	**31**

Thirty-one different unacceptable adjective noun collocational structures used 31 times.

One striking thing about the non-teacher norms adjective noun collocations is their degree of unintelligibility. Most of the non-teacher norms verb noun collocations may not result in breakdown of communication but this is not the case with the unacceptable adjective noun collocations particularly to an audience outside of the Nigerian communicative context. The 'strangeness' of the combinations highlights the difficulties L2 learners have with the production of incongruent adjective noun collocations. Only three out of the 21 structures can be identified as emanating from L1 negative transfer. If this is repeated across the four proficiency levels, that would suggest it is difficult for learners to transfer their L1 structures in the production of adjective noun collocations. I will now attempt to analyse these unacceptable collocations to have a better understanding of what goes on in the mind of the L2 learners in the production incongruent adjective noun collocation.

The expressions: 'lunatic + attitude/acts' and 'excretory + money' have their origin in Yoruba language. In Yoruba language, 'were' means lunatic (someone who is mentally ill) and 'iwa' means behaviour/attitude. 'Iwa were' therefore, literally means 'lunatic attitude/behaviour'. Similarly, 'acts' is 'ise', so 'ise were' is 'lunatic acts'. These Yoruba combinations are common, and the learners faced with conveying these ideas in English and not sure of the

acceptable way of expressing them resorted to their L1 knowledge to produce these infelicitous combinations. The acceptable collocation could have been 'aggressive + attitude/act' or 'belligerent + attitude'. The second expression: 'excretory + money' is a negative transfer from a Yoruba figurative expression for dirty money. By combining 'excretory' with money, the learners seem to be conveying, the concept of filthy money – ill-gotten wealth.

The other combinations are hard to explain; they are utterances that are, though grammatical, but cannot occur in correct natural English expressions. It seems because of the learners' increasing proficiency, they have become willing to take risks with their language production and in the process increasing the number of the infelicitous collocations they produced. Consider the following combinations: 'parental + image', 'nutritious + character', 'paramount + behaviour', 'restful + mind', 'abysmal + attitude', and 'unshakable + focus'. All of them are so strange that it is hard to figure out where the learners got the idea from. The only plausible explanation could be that they lack the awareness of restrictions on word combinations.

Meanwhile, the expressions 'junior + brother' and 'senior + brother' are very common in social interactions is Nigeria. The frequency data of the Corpus of Global Web-Based English shows that the expressions are extremely frequent in Nigerian and Ghanaian Englishes. However, these expressions are regarded as wrong in the classroom in Nigeria. This raises a few questions. If these expressions are widely used in social interactions in Nigeria as evidenced by frequency data from the Nigerian component of Corpus of Global Web-Based English, why then are they regarded as unacceptable by English language teachers? Since the expressions are widely used, should they not be accepted as features of Nigerian English? Why are these expressions regarded by the teachers as deviation and not variation? Does this mean the teachers are promoting conservative British English norms and standards by simply rejecting these expressions because they are not acceptable in the prestigious varieties? I will attempt to explain the linguistic justification for the rejection of these expressions despite them being widely used in social interactions in Nigerian speech community.

According to Lawal (2003:20), "a critical distinction between variation and deviation as two sociolinguistic/stylistic concepts is that whereas the speaker or writer constrains himself or herself within the structural limits of the language to select particular variant forms appropriate for his/her communicative needs, deviant forms, in stylistic terms, are reconstructed from the structural resources of the language to extend the frontiers of current usages". However, the issue with these expressions is that they are not a stylistic use of language where we could consider the reconstruction of 'younger brother' and 'older brother' as 'junior brother' and 'senior brother' respectively as forms to extend the frontiers of current usages. In this case, the teachers consider them as deviation from standard Nigerian English usage. The most plausible explanation for this could be because the concept of 'younger or older brother' semantically, is about the age (younger or older) and not necessarily a matter of being senior or junior in the literal sense of these words. So, it seems there is some linguistic rationale for deciding what counts as variation and what counts as deviation from acceptable norms in Nigerian English.

While much remains unknown about Nigerian English developmental stages from forming to norming, in some sense, Nigerian English teachers seem to be the promoters and drivers of norms and standards. While they regard certain expressions that are not in the British English as infelicitous, they also accept some expressions which are not in the British English but seem to be in consonance with the sociolinguistic reality of language use in Nigeria. But what is not clear is whether there is some arbitrariness in deciding what is unacceptable collocation and which collocation is in consonance with the sociolinguistic reality of language use in Nigeria and acceptable. I will address this further when discussing norms and standards in World Englishes in the discussion chapter. In the meantime, there is clearly a distinct variety of English in Nigeria – one of the emerging Englishes – which differs from the British English, particularly the lexico-semantics. The norm of this new English seems to be set by the English language teachers in Nigeria. This is the justification for using 'non-teacher norms collocation' for the collocational errors in this study instead of using 'non-native like' or 'deviant' for that would suggest the British English is the ultimate benchmark for deciding the correctness of collocations in Nigeria English.

8.3.2 Non-Teacher Norms Adjective Noun Collocational Structures in NILECORP-B2

This group of learners also produced fewer unacceptable adjective noun collocations in comparison to the unacceptable verb noun collocations they produced. There are 10 instances of non-teacher norms adjective noun collocation in the 130,559 words NILECORP-B2. In comparison to the errors in NILECORP-C1, this group of learners produced fewer unacceptable adjective noun collocations. As the data has shown, as proficiency increases, the production of collocations increases, and the instances of unacceptable collocations also increase. On the other hand, at lower proficiency, the production of collocations decreases as the learners used fewer formulaic expressions, and the instances of unacceptable collocations decrease. See table 8.8 below for details of all the non-teacher norms adjective collocations produced by this group of learners.

Table 8. 8 Non-Teacher Norm Adjective Noun Collocations in NILECORP-C1

Non-Teacher Norm Adjective Noun Collocations in NILECORP-C1	
Collocations	Frequency
tight + friend	5
truthful + friend	1
powerful + knowledge	1
malaria + medicine	1
headache + medicine	1
peaceful + humanity	1
Total	10

There are six different non-teacher norm adjective noun collocational structures in this table.

I will analyse the errors based on the possible source of the errors. Three out of the seven different collocational structures in this category can be attributable to negative L1 transfer. The expressions: ‘tight + friend’, ‘malaria + medicine’ and ‘headache + medicine’ seem to

have their origin in the Yoruba language. 'Tight friend' which was used five times in the corpus is a reference to intimate friendship (close friends). 'Ore' in Yoruba language means friend while 'timotimo' means very close. So, 'ore timotimo' means very close friend – as though something that is tightly closed. This is the origin of the expression 'tight + friend'. In Nigerian standard English, the acceptable collocation would be 'intimate friend' or 'bosom friend'. According to the frequency data from GloWbE, 'bosom friend' is not a common collocation in any of the prestigious varieties of English but frequently used in the emerging Englishes of Nigeria, Ghana, Kenya, and Asia. This will be discussed further in the discussion chapter of this study as one of the emerging themes – the existence of collocations in the New Englishes which are not in any of the prestigious varieties of Englishes (British, America).

The other two expressions: 'malaria + medicine' and 'headache + medicine' seem to be a direct translation of Yoruba to English. 'malaria + medicine' is translated from the Yoruba expression: 'ogun iba' – (ogun is medicine, iba is malaria) while 'headache + medicine' is from 'ogun efori' ('ogun' is medicine, 'efori' is headache). In this case, all these three non-teacher norms collocations are regarded as interlingual errors.

The other errors are intralingual and seem to be a result of the learners' lexical deficiency and lack of awareness of collocability of words. All these three combinations: 'truthful + friend', 'powerful + knowledge', and 'peaceful + humanity' are not natural in Nigerian English. While 'true + friend' is an acceptable collocation, 'truthful + friend' is not. In this case, this error seems to stem from the learners' lexical deficiency rather than lack of awareness of collocability of words. But the expressions: 'powerful + knowledge', and 'peaceful + humanity' which seem to be farther away from what is acceptable may be attributable to a combination of lexical deficiency and lack of awareness of collocability of words.

8.3.3 Non-Teacher Norms Adjective Noun Collocational Structures in NILECORP-B1

The third group produced only one non-teacher norms adjective noun collocational structure which was used four times in NILECORP-B1. Apart from the most proficient group, the

learners produced fewer adjective noun collocations. The unacceptable combination which they produced is: 'story + building'. It was caused by the learners' inability to discriminate the sounds of the co-occurring word(s) resulting in the substitution of collocating words with another word that sounds similar. They seem to have confused 'story' and 'storey' and thereby producing 'story + building' instead of 'storey + building'. This group's production of very few adjective noun collocations may be because of their avoidance of collocations.

8.3.4 Non-Teacher Norms Adjective Noun Collocational Structures in NILECORP-A2

The last learner group and the least proficient of the four learner groups also produced very few adjective noun collocations. This data has consistently shown that the least proficient groups produced fewer collocational errors than the most proficient group not because they have better knowledge of collocations but because they did not venture to produce as many collocations as the proficient group. They seem to use language 'safely' as opposed to the risk- taking proficient group – the C1 group. The errors come with the 'risky' use of language, but which also resulted in the production of many acceptable collocations. The least proficient groups which seem reluctant to take risk with the production of multiword units ended up producing few acceptable collocations and even fewer unacceptable collocational combinations.

In the 66,996 words NILECORP-A2, there are only three instances of non-teacher norms adjective noun collocations and 29 instances of acceptable adjective noun collocations. The three unacceptable collocational combinations are: 'unforgetful + holiday', 'break + money' and 'critical + accident'. In the first one, the learners combined an inexistent word (unforgetful) with holiday. The right collocation is 'unforgettable holiday'. The second one seems to be a direct translation from Yoruba. It refers to money to be spent during break while in school. Codeswitching is common in Nigeria, mixing English with Yoruba. The combination comes from a mixture of English and Yoruba – 'owo + break'. 'Owo' means money while using it with 'break' is a codeswitching expression which means money to be spent during break.

The learners' production of 'critical + accident' could be attributed to the use of synonymy as a lexical simplification strategy to get around the difficulty of producing the acceptable collocation. There are many adjectives that collocate with *accident* such as*: fatal, serious, tragic* and *ghastly*. (all these frequently collocate with accident in the Nigerian communicative context according to frequency data from GloWbE). All these collocates of accident refer to extremely serious event. It seems the learners consider 'critical' as synonymous to these adjectives hence producing 'critical + accident'. As indicated in the literature, which this study has also confirmed, learners resort to synonymy as a way of producing incongruent collocations. And the resultant collocation has always been a deviation from the acceptable norms of collocability.

In conclusion, a total of 268 non-teacher norms collocations were identified and analysed in this study. One hundred and twenty-eight of them representing 47.7% are attributable to negative L1 transfer while 140 representing 52.3% are caused by intralingual factors. The intralingual factors include synonymy, inability to discriminate sounds resulting in confusion, paraphrasing, and lexical deficiency. Two hundred and twenty instances of non-teacher norms verb + noun collocations were analysed within the context of their usage. Out of these figures, 115 representing 52.2% of all the unacceptable verb noun collocations produced by the four groups of learners are L1-induced while 105 representing 47.8% are intralingual errors. There are only 48 instances of non-teacher norms adjective collocations. Thirteen of them representing 27% are intralingual errors while the other 35 representing 73% are intralingual. The error analysis of the non-teacher norms collocations produced across all the four proficiency levels clearly shows incongruent collocations are problematic for the learners. It further shows an increase in the number of unacceptable collocations as the production of collocations increases. The more proficient learners produced more well-formed collocations and more non-teacher norms collocation. They also produced more L1-induced errors which seems to be an evidence of parasitic model of vocabulary acquisition. This will be discussed further in the main discussion chapter.

8.4 Discussion

The analysis of the non-teacher norms collocations produced by the four learner groups representing four different proficiency levels has revealed many things. The analysis was aimed at identifying the types of collocations are the most problematic for the Learners; the nature and causes of the collocational errors in the Learner Corpora; the similarities and differences in the error across proficiency levels and the proportion of collocational errors that are due to inter-lingual factors on the one hand and intra-lingual factors on the other hand. This discussion section aims to explain the findings within the immediate literature. I will discuss the findings under four themes namely: (1) incongruency the greatest cause of difficulty in L2 collocations production; (2) increase in the production of collocations means increase in the opportunity to make collocational errors; (3) L1 negative transfer is the biggest source of L2 collocational errors; and (4) evidence of parasitic model of vocabulary acquisition. This discussion is a prelude to the wider discussion chapter.

Various studies in the literature (Bahns & Eldaw, 1993; Nesselhauf, 2003; Yamashita & Jiang, 2010; Peters, 2015) have established that collocations that have no L1 equivalent are problematic for learners. In this thesis, all the learner groups have produced fewer incongruent collocations in comparison to the collocations that have L1 equivalent. This highlights the difficulty of producing incongruent collocations. The learners have instead produced more congruent collocations. It seems more convenient for them to produce language structures that are equivalent to their L1 while avoiding the structures that are incongruent. In this error analysis, all the non-teacher norm collocational structures identified in the four learner corpora are incongruent except one. This suggests that these L2 learners seem to rely heavily on their L1 in the production of L2 collocations. These findings lend credence to Bahns' (1993) call to focus on collocations that are incongruent to the learners' L1 in the language classroom as they are the most problematic. Meanwhile, as the findings on the effects of frequency of input on the production of collocations suggested in the previous chapter, if the incongruent collocations are frequent in the input the learners are exposed to, they become less problematic for learners to produce.

Beyond all this, the learners find polysemous collocations particularly problematic. As Macis and Schmitt (2016) note, polysemous could indicate different types of polysemy. But in this study, I use polysemous in the case of collocation to mean more than one literal or figurative meaning. When producing incongruent collocations, the learners, in their bilingual mental lexicon, seem to be mapping between words and concepts and figuring out which concept is expressed by a particular word (Pavlenko, 2009). This mapping seems problematic when the collocation involved is polysemous. Macis and Schmitt (2016: 50) identify three "meaning senses of collocations" namely: literal, figurative and duplex. In the first type, the literal meaning of the words forming the collocations are just added together (with semantically transparent meaning), but the second one has meanings that are not derivable from the co-occurring words. The duplex collocations, however, are polysemous. They use polysemous to mean having both literal and figurative meaning. The last two categories will probably be more problematic for learners.

Another theme that emerged in this chapter is that an increase in the production of collocations means increase in the opportunity to make collocational errors. My expectation was that the least proficient learners will produce more unacceptable collocations than the most proficient groups. But on the contrary, as proficiency increases, the production of acceptable collocations increases as well as an increase in the production of non-teacher norms collocations. What seems to have happened is that the least proficient learners are using language cautiously. Not willing to take risk with the language, they seem to avoid the production of collocational structures they are not sure of. What this means it that they produced fewer collocations which mean fewer opportunities to make collocational errors. But the most proficient groups on the other hand, buoyed by their increase in proficiency are more willing to take risk in their language use and adventurous with the production of collocations. In the process of production more collocations, it also provides an opportunity to produce more unacceptable collocations. This is not necessarily a bad thing as it means the learners are restructuring and recreating the language structure in their mental lexicon as well as testing hypothesis about the language. At some point in the acquisition process, it will result in increase in the production of acceptable collocations. The least proficient learners, however, feel safe with the production of congruent collocations and would not venture to produce unfamiliar collocations. This explains why there are fewer unacceptable collocations in their written texts.

As the more proficient learners restructure, formulate and text hypothesis on the production of more incongruent collocations, they seem to rely on a 'hypothesis of transferability' (Bahns, 1993: 61). The nature of the errors reveals a heavy reliance on the knowledge of their L1. The non-teacher norms collocations produced by the learners (and this is the same across all the four proficiency levels) are predominantly cause by L1 negative transfer. This is consistent with various findings in the literature (Farghal & Obiedat, 1995; Nesselhauf, 2003, 2005; Laufer & Waldman, 2011). All the four proficiency groups draw on their L1 metal lexicon to produce incongruent collocations. This supports the view of Wolter and Gyllstad (2011: 430) that "L1 may have considerable influence on the development of L2 collocational knowledge". But the negative effect of this is that the learners' reliance of their L1 means the production of unacceptable collocations. In this error analysis, most of the deviations consistently show attributes that are similar to lexical equivalents in Yoruba (the learners' L1). This evidence of L1-induced errors across the four proficiency levels seems to support the Parasitic Model of Vocabulary Acquisition (Hall, 1992). The parasitic model of vocabulary acquisition has as "its cornerstone the detection and exploitation of similarity between novel lexical input and prior lexical knowledge" (Hall & Ecke, 2003: 2).

The nature of the L1-induced errors seems to suggest 'parasitic learning strategy' (Hall, 1992) is their default mechanism for producing of unfamiliar collocations. The learners seem to process unfamiliar collocations based on similarity to their existing L1 knowledge. When producing incongruent collocations, it seems their "existing lexical representations … [are] activated and subsequently reconfigured" (Hall & Ecke, 2003: 2). This explains why L1-induced errors are predominant in the error analysis. Having completed the data analysis, the next chapter will focus on the discussion of all the findings.

Chapter Nine

Discussion

9.0 Introduction

The aim of this study has been two-fold: (1) to build a multi-level learner corpus of Nigerian English and (2) to investigate Nigerian English learners' use of collocations from World Englishes' perspective. The study started with a pilot study investigating the effect of immersion on Nigerian advanced adult speakers of English's knowledge of collocations. This pilot study led to a four-fold investigation of the collocational production and usage of Yoruba-speaking Nigerian learners of English which is, to the best of my knowledge, the most comprehensive study of collocations within the context of World Englishes. Firstly, the main study investigated the extent to which native and L2 learners use collocations in their written texts with a keen interest on the linguistic quality of the collocations they produced in terms of the span of the collocational string, and their structural and semantic properties. Secondly, it investigated the effects of frequency of and potential exposure to input in the learners' speech community on their production of collocations. Thirdly, it investigated the relationship between the production of collocations and proficiency (across proficiency levels). Finally, it identified, classified and analysed the collocations that deviate from the norms and standards of Nigerian English as opposed to the norms and standards of the prestigious varieties of English.

At the end of each of the analysis chapters (Chapters 5, 6, 7, & 8), I discussed the findings within the immediate literature. In this chapter, I will discuss the themes that emerged in the study within the wider literature on learner corpus research, L2 collocations and second language acquisition. This discussion is divided into two parts. The first part focuses on the themes that emerged from the first aim of this study - the building of the half a million words, first of its kind, Nigerian Learner Corpus of English (NILECORP) – a specialised learner corpus of young Yoruba-speaking Nigeria learners of English, and the concept of World Englishes. The discussion on learner corpus will focus on the assignment of proficiency levels to corpus texts and the value of more rigorous assignment of proficiency levels to

corpus texts in this study as well as the applications of NILECORP. I will conclude the first part of this chapter with a discussion on collocations in World Englishes and the question of norms and standards in the English language with specific focus on collocations in Nigerian English. The second part will then focus on the themes that emerged in the findings of this thesis. I discuss the collocational errors further considering the role of interlexical and intralexical factors in the production of collocations focusing on clang associations, frequency of input and congruency; and attempt to explain collocational links in L2 mental lexicon. I will attempt to explain the findings within Usage-based theory of language acquisition (Tomasello, 2003), Jiang's (2000) Model of Vocabulary Acquisition and Kroll and Stewart's (1994) Revised Hierarchical Model of bilingual language processing.

The entire chapter is, therefore, divided into five broad sections. The first section which is divided into three sub-sections discusses the design, development, assignment of proficiency levels, and the applications of NILECORP. The second section focuses on collocations in World Englishes, particularly on collocations that are in Nigerian English – one of the new varieties of Englishes, but which may not be in any of the prestigious varieties of English (British English/American English). It also examines the question of norms and the notion of error with specific focus on collocations in Nigerian English. I argue that the notion of standard in the English language can no longer be described as a homogenous phenomenon and as such, the application of exonormative standards would not be appropriate for Nigerian English.

In section three, which is the beginning of the second part of this discussion, I examine the differences in the production and usage of collocations by L2 learners and native speakers in relation to previous findings in the literature, particularly focusing on how my findings have widened our frontiers of knowledge in this area. The fourth section further explores the nature of the collocational errors produced by the learners and what they seem to reveal about their L2 mental lexicon. The fifth section discusses the principal findings of this study within the theoretical framework of Kroll and Stewart' (1994) Revised Hierarchical Model.

9.1 The Nigerian Learner Corpus of English (NILECORP)

Learner corpus has been described as a versatile source of data for second language acquisition research (Granger, 1998). Learner corpus has a wide-ranging application in applied linguistics as well as in all other language-related fields. The widespread use of corpora, not limited to learner corpus, has resulted in the development of many corpora in the recent years. But what seems to be missing is a specialised learner corpus designed within the concept of World Englishes. The learner corpus is a precursor to a bigger learner corpus Nigerian English which will include learners from other Nigerian L1s. In chapter four, I defined and described the population of the corpus, discussed the procedures for compiling it as well as the assignment of proficiency levels to its text. The discussion in this section is divided into three sub-sections. The discussion will focus on the assignment of proficiency levels to the corpus texts – an area that has not been well-researched in the learner corpus research literature, the common methods used for assigning proficiency levels in the literature, how the assignment of proficiency levels to NILECORP has contributed to this study, and the applications of NILECORP

9.1.1 Assignment of Proficiency Levels to Corpus Texts

The assignment of proficiency level to learner corpus texts is an important design criterion in computer learner corpora compilation but it is also somewhat a subjective notion as Granger (1998) rightly noted. A reliable proficiency level assignment of texts is essential for learner corpus research that compares learner groups. For instance, a corpus-based comparative study of Nigerian learners of English and Malaysian learners of English would need to know the proficiency level of the learners to ensure the comparison of the right learner groups. However, proficiency level, which Carlsen (2012) describes as a fuzzy variable in computer learner corpora, has not been the subject of much focus in learner corpus literature. Most of the learner corpus-based studies in the literature do report on whether their corpora are one-level or multi-level corpora (Guo, 2006; Kurosaki, 2013). A one-level corpus is a learner corpus that contains texts at one level of proficiency while a multi-level corpus contains texts at different levels of proficiency (Carlsen, 2012). But these levels of proficiency are not always clearly defined. Hulstijn et al. (2010: 16) lamenting the lack of reliable level assignment as a general problem in Second Language Acquisition research note that:

> "SLA [. . .] has frequently simply taken groups of learners at supposedly different levels of ability, conducted cross-sectional research and claimed that the results show development. Yet the levels have been woefully undefined, often crudely labelled 'intermediate' or 'advanced', or 'first and second year university students'—which means little if anything in developmental terms—and which cannot therefore be interpreted in any meaningful way".

In a similar vein, Carlsen (2012: 2) claims that "levels of proficiency are not always carefully defined, and the claims about proficiency levels are seldom supported by empirical evidence". She argues that the reliability of corpus-based research is jeopardized by the fuzziness of the proficiency variable. It is evident from most of the learner corpus studies in the literature that research agendas do not always grant enough attention to this issue. I will discuss the assignment of proficiency levels to NILECORP within the wider literature in Learner corpus Research. I will discuss the difficulty of assigning proficiency levels to corpus texts and the benefits of doing so.

Not clearly defining the proficiency level assigned to corpus texts calls into question the validity of claims made on such studies. It is important that the texts analysed in a learner corpus research are indeed representative of that particular proficiency level. But if a substantial part of the texts or in extreme cases, all the texts are not really at the knowledge and the ability in the use of the language assumed, this may invalidate any claim based on such data. Bachman (1990: 16) defines language proficiency as "the knowledge, competence, or ability in the use of a language, irrespective of how, where, or under what conditions it has been acquired". Sometimes language proficiency is referred to as language ability (Carlsen, 2012). A proficiency scale on the other hand as defined by the Council of Europe (CoE, 2001: 40) is "a series of ascending bands of proficiency. It may cover the whole conceptual range of learner proficiency, or it may just cover the range of proficiency relevant to the sector or institution concerned". If a researcher, for instance, assigns such labels as 'beginner', 'intermediate', or 'advanced' to the proficiency scale of a learner corpus text without a clear definition in terms of language descriptors, such assignment of levels may not yield meaningful information. This is one of the reasons why it is difficult to replicate certain studies in another context. The vague definition of the proficiency levels means it is impossible to determine the equivalent proficiency level in another context. For instance, how can we be sure that what a researcher refers to as 'intermediate' in a corpus-based study in Vietnam, for instance, is equivalent to what I label as 'intermediate' in a

corpus of Nigerian Learners of English. Such label does not say much about the linguistic ability of the learners in these two extremely different contexts – English is a foreign language in the former while English is a second language in the latter. English is, actually, the first language for some in Nigeria.

According to Carlsen (2012: 163), "a prerequisite for a reliable level assignment to texts should be an explicit definition of the theoretical construct underlying the assessment". As she rightly says, this construct validity of proficiency scales is of great significance to learner corpus research given the fact that a given proficiency scale is a valid representation of the underlying theoretical construct, and the way language proficiency is described at different levels in a learner corpus represents the stages of second language acquisition (Carlsen, 2012; Hulstijn, 2007). This is very important because a learner corpus with texts placed at proficiency levels according to a particular proficiency scale allows researchers to investigate the construct validity of that particular scale against empirical data (Carlsen, 2012). A reliable assignment of proficiency level to learner corpus texts means we can, as in the case of this study, investigate distinguishing features (in the production of L2 collocations) of each of the various levels of proficiency. Multi-level learner corpus texts, as Granger (2003: 8) rightly puts it, are "quasi-longitudinal" data because of the similarities between them and data collected from the same learners at different stages of their acquisition process. Such multi-level learner texts reliably placed at different proficiency levels enables us to empirically investigate the relation between proficiency scales and second language realities (Carlsen, 2012). All these highlight the benefits of multi-level learner corpora if proficiency levels are reliably assigned and clearly defined. Before discussing the method I used in the Nigerian Learner Corpus of English, let us first consider the methods which are commonly used to assign proficiency levels to learner corpus texts.

9.1.2 Methods of Assigning Proficiency Levels to Corpus Texts

The literature on learner corpus research reveals a multitude of different approaches to the assignment of proficiency levels to learner corpus texts (Tono, 2003; Carlsen, 2012). These different approaches can be categorised into two methods namely: learner-centred methods and text-centred methods (Carlsen, 2009; 2012). In the learner-centred methods, proficiency

levels are assigned to the texts based on the learners' characteristics and not the linguistic quality of the texts. These learner characteristics may be institutional status such as school year (class) or number of years the learner have been learning the language at an institution. An example of learner corpora that used this approach is The Uppsala Student English Corpus which is made up of essays written by Uppsala university students at three levels. The essays were written by the university students in the first term, second term and third term. However, there was no clear description of the linguistic ability of these students at the three different terms which could help to identify learners of equivalent proficiency in another context. Other characteristics which have been used to assign proficiency to learner corpus text include age of the learners, their total scores on a language test, or even the learners' teacher's opinion about their proficiency. An example of a learner corpus that used scores on a language test is the NICT Japanese Learner English which uses the scores of Standard Speaking Test to indicate the proficiency of each speaker's data. The clear definition of the proficiency levels will make it easy to analyse and compare the characteristic of interlanguage of each developmental stage and as well as compare it with learner corpus data with a clearly defined proficiency level. The Learner-centred methods of assigning proficiency level to corpus data seem to be the most widely used methods in the literature.

In NILECORP, proficiency levels were assigned to the texts using learner-centred method. Twenty-four English language teachers in Lagos who have taught the participants for up to five years, who have accessed the language ability of the students every term for up to five years determined their proficiency levels. Based on their knowledge of the participants' language performance, they situated the learners' language ability within the Common European Framework of Reference for Languages (CEFR) matching the learners' linguistic ability with the corresponding language descriptors for the six proficiency levels in the CEFR as discussed earlier in chapter four. This seems to be the first time such a method was used in the literature using the learners' teacher to situate their proficiency within the CEFR. The other study which used CEFR proficiency grid but not through the learners' teachers' opinion is Carlsen (2012) who linked the *Andrespråks-korpus* (ASK) – a learner corpus of Norwegian as a second language to the Common European Framework of Reference for Languages. One benefit of this is that, researchers in other parts of the world who might not have understood the linguistic ability of the learners if I had used such labels as 'second year', 'third year', or 'fourth year' high school students in Nigeria may be able to check the CEFR

language descriptors and use this study for a comparative study in other contexts. While I acknowledge there may be some elements of subjectivity in the teachers' opinion on the learners' proficiency, their analysis of the corpus data clearly shows distinguishing linguistic features that characterise the interlanguage across the four proficiency levels. This may be considered as the validity of the method. However, the shortcoming of this method is that it does not account for individual proficiency of the participants. This method may not be appropriate if the aim of the research is to account for individual linguistic ability of the participants. But in this study, the aim is to look at the collective knowledge and use of collocations and to understand the effect of frequency of exposure to the target structure in the learners' speech community as well as the effect of their L1. This method of assigning proficiency to texts seems most practicable when compiling relatively large corpus where the producers of the texts can be identified. Besides, their teachers must know the learners long enough to provide reliable data on their linguistic ability.

Carlsen's (2012) Corpus Texts Levels Assignment table which I have reproduced below with slight modification to include some elements from Atkin et al (1992) clearly shows the difference between learner-centred and text-centred methods.

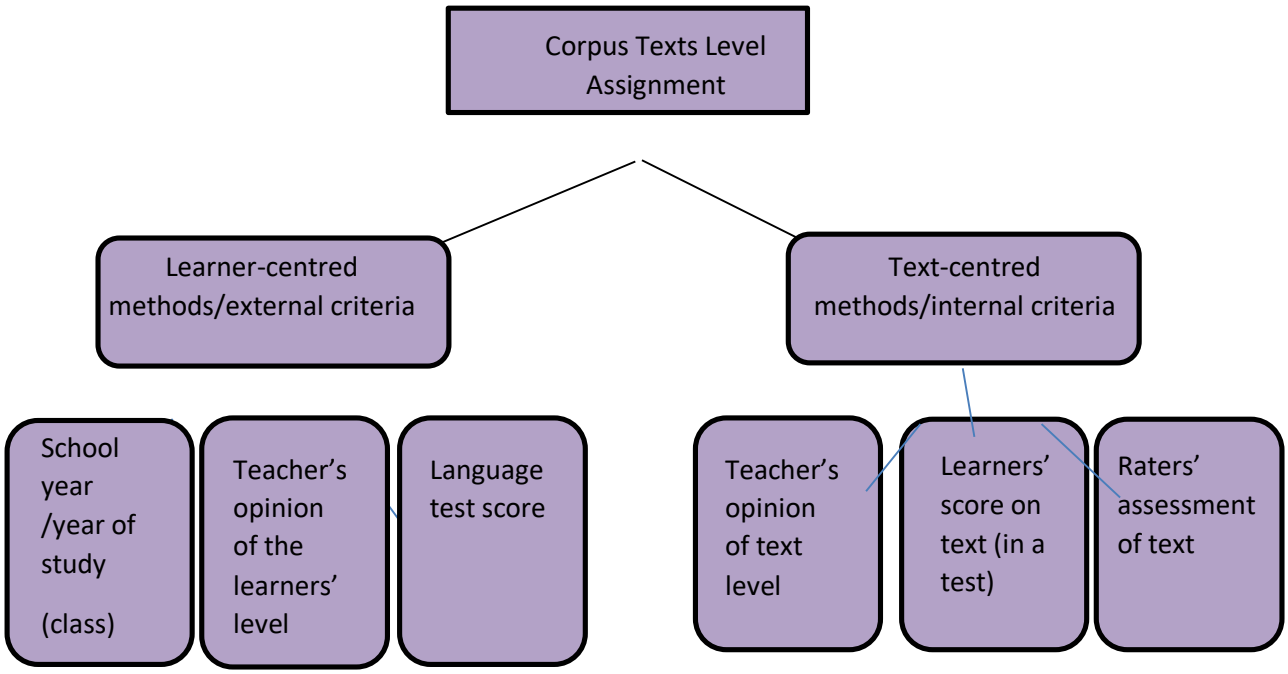


Figure 1: Methods of proficiency level assignment to learner corpus texts (after Carlsen, 2012: 166)

Atkin et al's (1992:5) distinction between 'external and internal criteria' for constructing a corpus for linguistic analysis is similar to the distinction between learner-centred and text-centred methods. The internal criteria which correspond to text-centred methods are essentially linguistic – the classification of text according to its linguistic characteristics. The external criteria, on the other hand, are non-linguistic. They are non-linguistic attributes which are considered relevant to the description of the language population where the learner corpus texts come from. Just like the learner-centred methods, external criteria for assigning proficiency level on the text can be determined without reading the text in question.

Atkin et al (1992) conclude that a corpus selected entirely on external criteria would be liable to miss significant variation among texts since the assignment of proficiency level is not motivated by textual factors. This conclusion opens a whole array of criticism of the learner-centred methods (external criteria). One of such criticism comes from Tono (2003: 801) who argues that: "selection based upon external criteria such as school year or age does not necessarily ensure that the subjects are comparable in terms of language proficiency". He uses the case of Japanese-speaking EFL learners group in comparison to learners from European countries. Although their learner profile fulfilled all the criteria, their proficiency levels, however, are so markedly lower than those from European countries. What this means is that learner corpus text from Japan or China, for instance, labelled as 'second-year university English-majors' may not be equivalent to similar texts from Nigeria or Netherlands in terms of their linguistic characteristics (proficiency level). Do all these now invalidate learner-centred methods of assignment proficiency levels to corpus text? The issue here is not necessarily the label: 'second-year university English-majors' but the context. The proficiency level of second year university English majors students in Japan where in English is an international language may not be the same with second year Nigerian university English majors students where English is a second langue. One plausible way of addressing the problem of levels assignment to corpus text would be a clear definition of the levels assigned to corpus texts in terms of language descriptors which is what I did in the assignment of proficiency to NILECORP. A label such as 'second-year university English-majors' as I have been emphasizing does not, even in the vaguest way, say what language abilities the learners have.

The other methods used for assigning proficiency to corpus texts in the literature are text-centred methods. In text-centred methods, proficiency levels are assigned to corpus texts based on the linguistic quality of the texts irrespective of the learners' characteristics and their other language skills (Carlsen, 2009). As Carlsen (2009) puts it, proficiency level can be assigned to the texts based on the learners' teacher's opinion about their texts, scores of a written essay (or oral part for a spoken corpus) in a standardised language test, or similar text assessment rated by various experts to ensure validity of decisions. Using text-centred methods or internal criteria to assign levels to corpus texts is by no means less problematic. The text-centred methods will require analysis of a range of linguistic features of the texts which will contribute to its characterisation in terms of internal evidence to determine the proficiency level (Atkin et al, 1992). As Marchand and Akutsu (2015) rightly said, in order to make the use of text-centred methods to assigning proficient to corpus texts practical and easy, consideration must be given to the length of the corpus texts, and the tools/the criteria for assessing the texts must not be overly taxing on the raters. It will require great effort to go through the learners' texts in a big corpus. Perhaps this explains why text-centred methods are not frequently used in the literature. Whatever the case, a corpus text selected entirely based on learner-centred methods would be liable to miss significant variation among texts so also a "corpus selected entirely on internal criteria [text-centred method] would yield no information about the relation between language and its context of situation" (Atkin et al, 1992: 8). What would matter most is a clear definition of the linguistic ability that the texts represent.

As stated earlier in the literature review chapter that the various studies on collocations in Nigeria did not clearly define the linguistic ability which the texts they analysed in their research represent. And the proficiency they assigned to their texts did not have any empirical support. However, by going through the rigour of assigning proficiency levels to the corpus texts instead of crudely labelling the texts as second or third year high school students in Nigeria, this study has provided findings that can be interpreted in developmental terms. By using an internationally recognised proficiency levels, the findings of this study can be compared with learners with similar proficiency elsewhere.

9.1.3 The Applications of NILECORP

NILECORP is obviously a versatile tool for linguistic inquiry not just into the distinguishing features of Nigerian English but also for comparative corpus-based analysis of varieties of English. It also has various pedagogic applications. The discussion in this sub-section is focused on the main applications of the Nigerian Learner Corpus of English namely: material design, pedagogic lexicography, teaching methodology and learner corpus research.

Learner corpus data has "tremendous potential…to inform pedagogical tools and methods" (Granger, 2017:345). The Nigerian Learner Corpus of English (NILECORP) has various pedagogic applications. Three aspects of language education which may benefit most from the corpus-informed insights that NILECORP may offer are: material design, pedagogic lexicography and teaching methodology. It can help to design corpus-informed in-house teaching materials which may be L1-specific rather than generic. Such materials could address L1-induced lexico-grammatical difficulties. This could be of a greater pedagogical significance than the global ELT coursebooks which are far removed from the Nigerian local learning context. The ELT coursebooks in Nigeria are a combination of the global ELT coursebooks and locally made coursebooks. However, most of the locally made coursebooks are not corpus-based but based on intuition and experimental data.

Materials designed based on the teacher's intuition relies on an individual's intuition, whereas corpus data offers a pool of "intuitions of a great numbers of speakers" (McEnery & Xiao, 2011: 364). Corpus data can complement or even refute the intuition of individual teachers which may not always be reliable (McEnery & Xiao, 2011). As for experimental data, they "may contain artificial interlanguage forms" (Granger, 2008: 337), but learner corpus offers authentic data which demonstrates how the Nigerian Yoruba-speaking learners of English use the language "when they are primarily engaged in message construction" (Ellis & Barkhuizen, 2005 cited in Granger, 2008: 337). Corpus-based or corpus-informed pedagogic materials can address what intuition and experimental data-based pedagogic materials, potentially, miss out. This means language teaching professionals in Nigeria can use the corpus data to design pedagogic materials that address the specific lexical and grammatical needs of the learners as revealed by the learner corpus data. For instance, the

analysis of NILECORP reveals that collocational deficiency is pervasive as well as revealing the most problematic collocations for Yoruba-speaking learners of English. It also reveals the most frequent errors and the causes of the errors, e.g. L1 interference. This insight can inform the teacher's decision which may result in the design of tailor-made pedagogic materials to address learners' specific needs. Alternatively, these corpus-based pedagogic materials could be used to supplement the existing teaching materials.

The application of learner corpus data to the design of pedagogic materials has been widely acknowledged in the literature (Tono, 2003; Nesselhauf, 2004; Granger, 2008, 2012; McEnery & Xiao, 2011; Xu, 2016). Nesselhauf (2004) stresses that one of the greatest potentials of learner corpus is that it can be used to improve pedagogic materials. This view was supported by Granger (2012: 22) who points out that the "fields that have benefited most from learner corpus insights are lexicography, courseware and language assessment". There are already many corpus-based English language coursebooks in the market. However, none of the corpus-based/corpus-informed English coursebooks available in Nigeria have their insight from Nigerian English corpus. But the existence of these coursebooks is a testimony to the veracity of corpus-based pedagogic materials. So, using insights from NILECORP to design coursebooks mean addressing context-specific needs of the Nigerian learners and probably learners in other similar contexts.

NILECORP could also be helpful in the areas of pedagogic lexicography if used in conjunction with a corpus of advanced speakers of Nigerian English. I use pedagogic lexicography, to mean all dictionaries conceived for learners of a second or foreign language (Tarp, 2011). There have been a few corpus-based dictionaries of collocations, e.g. *Oxford Collocations Dictionary for Students of English.* While the existing dictionaries reflect the use of collocations in authentic context, they are based on the prestigious varieties of English excluding collocations in the emerging Englishes like Nigerian English. While there is a dictionary of Nigerian English Usage (Igboanusi, 2002), it does not necessarily focus on collocations in Nigerian English. A dictionary of Nigerian English with a focus on collocations could benefit from insights from NILECORP by incorporating error notes generated on the basis of the corpus to help Nigerian leaners avoid making common mistakes. As Granger (2008: 344) points out, "these notes are a clear added value for

dictionary users as they draw their attention to very frequent errors". What this means is that only a Nigerian English corpus is in a better position to reveal the frequent errors peculiar to the Nigerian learners of English. While I am not saying the global corpus-based dictionaries are irrelevant to Nigerian learners, when it comes to the specificity of L2 English common and frequent errors in the Nigerian context, only a Nigerian English corpus-informed dictionary can address them properly.

On the pedagogic applications of NILECORP, I want to discuss its applications to teaching methodology. The focus will be on the application of NILECORP in conjunction with advanced speakers' corpus to Corpus-driven Instruction (DDI) and Data-driven Learning (DDL) with a caveat that the appropriateness of using a learner corpus for both DDL and DDI is dependent on the learner objectives. Corpus-driven instruction is the use of "corpus-based reference grammars, textbooks, and dictionaries that include attested language samples instead of invented examples" (Vyatkina, 2015: 1) in the teaching and learning process. This is an indirect application corpus data to language teaching. Language pedagogy could benefit from an expanded corpus-driven instruction which will include the application of raw learner corpus data to classroom instruction. This is a kind of teaching method in which the teacher uses corpus-based information on the interlanguage of Yoruba-speaking leaners of English to improve instruction. This data which contain examples of frequent errors, among other things, will inform the choice of class activities, the examples of the target structures to be used in class and the whole pattern of the instruction. Essentially, the teacher relies on corpus information to inform her teaching and learning. In this way, the teaching is less subjective, more objective and less intuition-based. This will provide English language teachers the ability to be more responsive to learners' specific needs particularly focusing on L1-induced factors in language learning.

NILECORP could also be applied to Data-driven Learning but with obvious limitations in the Nigerian context. Data-driven Learning is "the use in the classroom of computer-generated concordances to get students to explore regularities or patterning in the target language, and the development of activities and exercises based on concordance output" (Johns & King, 1991: iii). The main thrust of the method is for learners to discover the target structure "from multiple occurrences in context, augmented with lists and charts of frequencies, collocates,

wordsketches…" (Boulton, 2017: 6). Corpus data which will provide examples of 'multiple occurrences in context' "can provide enough evidence and stimuli for the learners to arrive at developmentally-appropriate generalisation" (Bernardini, 2004:17). With the appropriate software, NILECORP can provide concordances for learners to "explore regularities or patterning in the target language" (Johns & King, 1991: iii). This has great potential as it means Nigerian learners can explore, for instance, collocations in Nigerian English. DDL has not made its way into the mainstream language teaching methods in Nigeria. One obvious reason for this is the lack of the required technology for this method. Boulton (2017: 6) argues that "technological advances have made DDL faster, simpler, more intuitive, prettier, more accessible…" This could be the case in developed countries but certainly not true about the Nigerian context.

Another apparent downside of this teaching method, particularly in the Nigerian context is that its success is hinged on "the learner's ability to find answers to their questions by using software to access large collections of authentic texts relevant to their needs, as opposed to asking teachers or consulting ready-made reference materials" (Boulton, 2017: 1). By putting the learners in the driver's seat, the success of the learning process will only be commensurate to the learners' ability to know what to query in the first instance. Even when the learners know what to query, there is still the problem of "formulating the question as a query that the software can understand, and then interpreting the results" Boulton, 2017: 7). Having said that, DDL may be very helpful for relatively advanced learners who know what to query, able to formulate their questions in a way the software can understand and interpret the results.

Finally, NILECORP could be used for various studies on error analysis, the quantitative differences between the interlanguage of various varieties of English, the description of the features of the interlanguage in its entirety, and the application of learner corpora-based research to language teaching methodology and materials design. The learner corpus is suitable for corpus-based error analysis as well as research on the development and evaluation of automatic detection of errors and tagging. As learner corpora offer examples of authentic language use, NILECORP could be a useful tool for researchers who are interested in the quantitative differences in the use of certain syntactic, lexical and discoursal features

between the interlanguage of various varieties of English. Such corpus-based studies will provide data on whether certain learners use particular linguistic features more frequently or less frequently than others.

NILECORP, being a multilevel learner corpus representing four proficiency levels, researchers could exploit it to describe the overall characteristics of the interlanguage either at a fixed stage or at different developmental stages – in this case, four developmental stages. The learner corpus is also useful for those who are interested in the pedagogical applications of the results of analyses of learner data to improve various aspects of language pedagogy. This may be very relevant to language teacher education in Nigeria. There are, to the best of my knowledge, no studies on relating the findings from learner corpora to actual classroom practice.

9.2 Collocations in World Englishes

There has been an increasing interest in L2 collocations research. This interest could be attributable to our increasing awareness of what a problematic linguistic phenomenon collocations are in second language acquisition and the availability of both small and large corpora as well as the available corpus analysis software. The focus of the studies in the literature has been on two types of collocations: lexical collocations and grammatical collocations (Benson, Benson & Ilson, 1986). Most of the existing studies seem to focus more on lexical collocations. Some of these studies delimit their investigation to one type of lexical collocation (e.g. Farghal & Obiedat, 1985; Bahns & Eldaw, 1993; Nesselhauf, 2003, 2005; Holtz, 2007; Siyanova & Schmitt, 2008; Laufer & Waldman, 2011) while some have a range of lexical collocations (e.g. Groom, 2009; Yamashita & Jiang, 2010). All these studies point to the fact that collocations are both pervasive in the English language and difficult for learners including advanced speakers of English as a second language. This is the main thrust of my finding as I have said earlier. However, this current study is distinct in many ways. It conceives and operationalizes collocations within the concept of World Englishes; it investigates the structural and semantic properties of collocations in learner corpus versus native corpus; and has a wider scope than most of the studies in the existing literature.

This study brings to the fore a new perspective on the conception of collocations, a perspective that advocates for the consideration of the learners' speech community (the variety of English spoken in the country) in defining the concept of collocations. English language can no longer be considered as a single monolithic entity. On the contrary, there are now new Englishes which as I said earlier, are still in communion with their ancestral home but altered to suit their new environment. These new Englishes have been variously referred to in the literature as "institutionalized non-native varieties of English" (Lowenberg, 1986), "world English" (Kachru, 1992), "indigenized Englishes" (Mufwene, 2015), "New Englishes" (Platt, Weber & Ho, 1982), "extraterritorial English" (Lass, 1987), "postcolonial Englishes" (Schneider, 2007), etc. As Crystal (2003: 146) puts it, "most adaptation in a New English relates to vocabulary, in the form of new words …, word-formations, word-meanings, collocations and idiomatic phrases". Considering the variety of the English which is spoken in the learners/users' speech community in defining collocations will account for the collocations in such variety of English which may not necessarily be in any of the prestigious varieties of English.

To better explain this, I will provide some examples of collocations which are frequently used in Nigerian English but may not be regarded as collocation in native English because they hardly co-occur. The verb "proffer" for instance, frequently co-occurs with the noun "solution" in Nigerian English forming the verb noun collocation: "proffer + solution" which means to offer solutions – and this collocation is apparently exclusive to Nigerian English. Another example (extracted from NILECORP-C1) is *"social + miscreant"*. This adjective noun collocation which means someone who behaves badly in public places in big cities is frequently used in Nigerian and Ghanaian English according to frequency data from the Corpus of Web-Based Global English (GloWbE). And again, we have the adjective *"nonchalant"* which, according frequency data from GloWbE, co-occurs frequently with the noun *"attitude"* in Nigerian English than in any other varieties of Englishes including the native Englishes. All these are examples of collocations in Nigerian English, one of the emerging new Englishes. There are probably thousands of such collocations in Nigerian English which are part of the distinguishing features of that variety of English, but which may not be in any of the prestigious varieties of English. But this could not have been limited to Nigerian English as there are many other new varieties of English in Africa and Asia. These emergent varieties of English are mainly in the former colonies of the United Kingdom.

They are part of Kachru's (1992) outer circle English. These new Englishes have developed distinctive and stable lexical (including collocations), syntactic, phonetic and phonological characteristics. These varieties of Englishes are spoken by many either as a first language (L1) or as a second language (L2).

With the existence of new Englishes and new collocations which may not exist in the prestigious varieties of English, the use of such expression as 'non-native-like' and strictly assessing learners' knowledge of collocations on the basis of norms and standards of the prestigious varieties of English is becoming problematic. Also problematic is the notion of 'native speaker'. The global use of English and the fact that many people now speak English as their L1 further problematize the notion of native speaker. There seems to be a gap in the existing literature on the existence of new collocations in world Englishes and how this may affect our judgement of what counts as acceptable and unacceptable collocations. This current study, to the best of my knowledge, is the first to conduct a largescale corpus-based study of collocations. While a native corpus has been used as some reference corpus, all instances of collocations which acceptable in Nigerian English but which not in the reference corpus are included in the study. All such collocations are credited to the learners and not regarded as evidence of collocational deficiency even though such collocation may be considered by speakers of some of the prestigious varieties of English as infelicitous. Any study of L2 collocations of Nigerian speakers of English or speakers of any of the other varieties of World Englishes that does not take cognisance of the existence of the collocations that may be peculiar that variety would not produce an accurate understanding of their collocation knowledge.

9. 2.1 Collocations in World Englishes: the question of Norms and the Notion of Error

There is a consensus in the literature on the multiplicity of the English language. What this means it that the notion of standard in the English language can no longer be described as a homogenous phenomenon and as such, the application of exonormative standards would not be appropriate. An endonormative standard will account for the various features of the new Englishes as used in diverse sociolinguistic contexts around the world. Deciding what counts

as acceptable innovative use of language and what counts as error in World Englishes may be problematic – and this includes collocations in World Englishes. As it is, these new varieties of Englishes are still evolving. If there was a dictionary of Nigerian English collocations as we have it in the prestigious varieties of English, that would have provided some form of codification. As Bamgbose (1998: 4) puts it, "once a usage or innovation enters the dictionary as correct and acceptable usage, its status as a regular form is assured". But in the absence of such codification, deciding which collocations in Nigerian English is acceptable and which one is not is still considerably hazy. This section of the discussion addresses this conundrum.

Though the existence of new Englishes is widely acknowledged, the conflict between using exonormative standards and endonormative standards still exists. Jowitt (1991: 47), describing Nigerian English observes that "the usage of every Nigerian user of English is a mixture of Standard forms and Popular Nigerian forms, which are in turn composed of errors and variants". While he acknowledges the existence of a legitimate variant which he describes as 'Popular Nigerian forms', he however still contrasts it with 'Standard forms' by which he was obviously referring to British English. What he seems to ignore is that, as Dürmüller (2008: 241) puts it: "in the profile of these new varieties, particularities can be detected in pronunciation, spelling, lexicon, grammar, semantics (word, phrase and text meanings), and in pragmatics which make them differ, not only from each other, but also from the established standard varieties". In view of this, contrasting Nigerian English with British English premised on exonormative standards means delegitimising the Nigerian variant. Using exonormative standards will regard all the innovative use of language and other collocational expressions which are reflections of the sociolinguistic reality of language use in Nigeria as errors just because such expressions are not in British English.

The question then is who determines the endonormative standards for the new Englishes? Who should be the gatekeepers of Nigerian English standards? One of the most plausible answers would be linguists, policymakers and English language teachers who are always the gatekeepers and main transmitter of norms (Schneider, 2007). Looking at this as a researcher and member of Nigerian English Language Teachers' community, teachers occupy a pivotal position to determine what counts as an innovative use of the English language and hence acceptable and what counts as a deviation from acceptable language use in the Nigerian

context. Much of what should be regarded as acceptable Nigerian collocation is more of a question of whether such expressions are widely used and accepted in the Nigerian speech community. Just as Carter and McCarthy (2006: 5) rightly said, the "issues of acceptability are never far from the surface when there is reference to what is standard in grammar or in language use in general". This is closely related to Banjo's (1993) argument that an endonormative model for Nigerian English must pass two tests namely: local acceptability and international intelligibility. One way of deciding whether the Nigerian collocations are widely used and accepted is to check how frequently they are used in corpus of Nigerian English. As there is no codification of standard Nigerian English for now, it should suffice to use an endonormative standard based on acceptability in Nigeria as determined by English language teachers' judgement and frequent use in Nigerian corpus of (advanced speakers of) English. I acknowledge this is subjective to some extent, but it remains the most plausible solution in the absence of codification.

While this area needs much empirical research, the thrust of my argument is that the English language in Nigeria (as well as in other contexts where there exists new varieties of English) "has been acculturated and transmitted to release multiple characteristics deviant from its mother in the Inner Circle … obsolete ELT paradigm, that is based on the ideology that native speakers are the authority of the language, needs to be replaced by a newer paradigm that relates language classroom to the world and takes into account local adaptation and appropriation" (Jindapitak and Teo, 2013:197). And as such, it is not appropriate to use the norms and standards of the prestigious varieties of English as a benchmark for deciding what is correct and what is not in Nigerian English. A paradigm shift is necessary and L2 collocations research needs to reflect this, at least in context where there is an emerging variety of English.

9.3 Collocations in Learner Corpus versus Native Corpus

This section which is the beginning of the second part of the discussion chapter elaborates further on the discussion of the findings of the comparative analysis of the collocations produced in NILECORP-C1 and LOCNESS. This is a continuation of the discussion I started

at the end of chapter five. By way of a reminder, NILECORP-C1 is the most advanced of the four learner groups while LOCNESS is the native English corpus. As I have said earlier, the second part of this discussion chapter focuses on the themes that emerged from the findings of the study while the first part focused on the themes that emerged as biproducts of this study. It aims to discuss the findings within the existing literature on comparative analysis of collocations in native and non-native corpus.

As I have said earlier, there has been a growing body of literature comparing native speakers' use of collocations with non-native speakers' use of collocations, comparing L2 learners' use of collocations across various proficiency levels, effect of exposure to input and a host of other variables on collocational production, (Bahns & Eldaw, 1993; Bahns, 1993; Nesselhauf, 2003, 2005; Siyanova and Schmitt, 2008; Groom, 2009; Durrant & Schmitt, 2009; Laufer & Waldman, 2011; Demir, 2017). The overwhelming consensus in the literature is that L2 speakers, regardless of their proficiency level, deviate from native speaker norms in their production of collocations – all the existing L2 collocational studies are benchmarked against native speaker norms which this study is challenging. And the degree of the deviation varies across proficiency levels and the context of learning (whether learner lives in the target language context or not).

Durrant and Schmitt (2009) investigate the extent to which native and non-native writers make use of high-frequency collocations with a focus on strong collocations in comparison to native speaker norms. They conclude that "non-native writers rely heavily on high-frequency collocations, but that they underuse less frequent, strongly associated collocations (items which are probably highly salient for native speakers)" (ibid: 157). In a similar study, Demir (2017: 84) who compares the use of collocations in texts produced by native English authors and Turkish L2 English authors concludes "there are robust differences between native and non-native writers in terms of using lexical collocations … [and a] close relation between nativity of the authors and the number of collocation[s] which were used". He further points out that it is "highly apparent that native authors used much more collocations than Turkish authors" (ibid: 84). Laufer and Waldman (2011) also compared the production of L2 collocations in a multilevel learner corpus representing three proficiency levels (basic, intermediate and advanced) with native speaker corpus. Their results show that the learners at all the proficiency levels produced far fewer collocations in comparison with the native speakers. The number of collocations in the learner corpus only increased at the advanced

level. Their data shows interlingual errors are persistent even at advanced levels of proficiency.

All these studies seem to confirm the common position in the literature that collocational deficiency is pervasive even among advanced learners of English (Granger, 1998; Nesselhauf, 2003, 2005). The findings of Demir (2017 and Laufer and Waldman (2011) show that non-native speakers produced fewer collocations than their native speaker counterpart. However, my findings are quite the opposite. Based purely on the frequency of the instances of collocations regardless of how many times a particular collocational structure is repeated, the L2 learners produced slightly more collocations in their text than the native speakers. But if we consider the numbers of different collocational structures produced, the native speakers produced slightly more collocations than the L2 learners. This raises a number of issues which I am going to highlight in this discussion in an attempt to account for these apparent contradictory findings. Most of the comparative studies in the literature have often concluded by saying, for instance, that non-native speakers produced fewer collocations than native speakers. Can such conclusions be taken across the board to mean that in all instances, non-native speakers produce fewer collocations than native speakers?

To have a better understanding of non-native speakers' usage of collocations in relation to native speakers, we will have to clearly define the 'non-natives' we are comparing with the native speakers. Various factors can affect the acquisition of a second language (including the production of L2 collocations) and these include: linguistic distance between the L1 and L2, the learners' proficiency level in the L2, the learning context among other things (Walqui, 2000; Collentine & Freed, 2004; Montero, Serrano & Llanes, 2017). In the case of the L2 learners in this study, they speak Yoruba as L1 – a language that is linguistically distant from English, and their proficiency is equivalent to the CEFR C1 level. All the learners live in an English as a second language context where there they are exposed to the target language frequently. All these factors might have impacted the acquisition process in some ways. With all these variables in mind, this study shows, contrary to Laufer and Waldman (2011) and Demir (2017), that relatively advanced learners (CEFR – C1 equivalent) of English from an English as a second language context where the learners have frequent exposure to the input outside the classroom, produced more collocations than the native speakers, albeit, a

narrower range of collocations. This study uses the same native speaker corpus (LOCNESS) which Laufer and Waldman (2011) used.

Another reason for the opposing findings could be the proficiency level of the learners involved in the study. Demir (2017) does not clearly define the non-native authors he was comparing with native authors in terms of their English language proficiency. He only describes them as 'Turkish authors' which as Hulstijn et al. (2010: 16) rightly points out "means little if anything in developmental terms—and which cannot therefore be interpreted in any meaningful way". This somehow seems to cast some doubt on his findings. Without the clear definition of the proficiency of his 'Turkish authors', it is difficult to compare his findings with any other study comparing native and non-native speakers' use of collocation. It is important to establish the linguistic ability (the proficiency level) of the Turkish authors to compare his findings with the findings of studies that investigate other non-natives at the same proficiency level. Laufer and Waldman (2011) describe the proficiency of the learners in their study as basic, intermediate and advanced. They called the L2 learners at the level of 9th and 10th graders "basic," the ones at the level of 11th and 12th graders "intermediate," and the college and university students "advanced". Even then it is still difficult to know what these means for comparative purpose. This further highlights how problematic the assignment and description of learners' proficiency levels have been in learner corpus research including many of the studies that compare native and non-native speakers' use of collocations.

This current study has shown that the learners used more of the collocations that are frequently used in Nigeria and they used fewer of the less frequently used in Nigeria according frequency data from the Nigerian component of GloWbE. These findings seem to confirm Durrant and Schmitt's (2009) findings. The findings suggest that learners are more likely to acquire and use collocations that are frequently used in their speech community (learning context). This points us to the usage-based model of language acquisition (Tomasello, 2003). According to the usage-based model of language acquisition, frequency of occurrence and co-occurrence of linguistic forms in the input the learners are exposed to are the main determinants of the acquisition of formulas (Barlow and Kemmer, 2000). Frequency and linguistic experience are very crucial to a usage-based approach.

The correlation between the collocations the learners produced in the learner corpus and the frequently used collocations in the Nigerian component of the GloWbE could be the result of the frequency of the co-occurrence of linguistic forms in the input they are exposed to in Nigeria. The learners might have learned these frequent collocational structures through "intention-reading" and "pattern-finding" in their linguistic experience (Tomasello, 2009: 69). Intention-reading, as Tomasello (2009: 69 - 70) puts it, "is what children must do to discern the goals or intentions of mature speakers when they use linguistic conventions to achieve social ends, and thereby to learn these conventions from them culturally". Pattern-finding, on the other hand, "is what children must do to go productively beyond the individual utterances they hear people using around them to create abstract linguistic schemas or constructions" (ibid: 70). This exemplar-based model explains child's L1 acquisition process which is based on frequency-based analysis of memorised patterns, but it could plausibly help to explain how frequency of and exposure to input affect L2 acquisition.

The frequent use of certain co-occurring patterns in Nigeria provides the learners frequent exposure to multiple instances of collocations which means more opportunity for intention-reading and pattern-finding. Also, the less frequently used co-occurring patterns in Nigerian means less exposure to such patterns and fewer chances for intention-reading and pattern-finding. This could explain why the learners used more of the frequently used co-occurring patterns in Nigeria and fewer of the less frequently used patterns. What this means in simple terms is that the higher the frequency of the co-occurring patterns in the input the greater the chance of acquisition, the less frequent the patterns are in the input, the less the chance of acquisition. This seems consistent with various studies that show a strong relationship between frequency of exposure and language acquisition and processing (Ellis, 2002; Durrant and Doherty, 2010; Kim and Kim, 2012; Walter and Gyllstad, 2013; Gonzalez and Schmitt, 2015).

9.3.1 Semantic Properties of L2 Collocations

The discussion in this section centres on the semantic properties of collocations produced by the learners in comparison to the collocations produced by the native speakers. One aspect in

which our knowledge seems severely limited is the semantic properties of collocations produced by L2 learners – how L2 learners use collocations with figurative meaning. Most of the existing comparative studies have focused on the quantity of collocations produced by L2 learners in comparison to native speakers rather than the linguistic quality of the collocations produced. This current study has shown no significant difference, in quantitative terms, between the collocations produced by the most advanced group of learners and the native speakers. However, there is a difference in the structural and semantic properties of the collocation produced by the leaners and native speakers. The native speakers produced far more collocations with additional meanings than the L2 learners. The difference in the semantic properties of the collocations produced by the native speakers and the learners is very pronounced. Equally pronounced is the difference in the semantic properties of collocations produced across the four proficiency groups. This study shows a link between the learners' production of semantically burdensome (referring to the semantic properties) collocations and their L2 English proficiency and age. Just as a reminder, the semantically burdensome collocations are collocations which are "imbued with a bewildering range of connotative and associative meanings" (Phillip, 2011: 26). The more proficient learners in this study who are also the oldest learner group produced more of these figurative collocations than the less proficient groups who are younger. The oldest of the four groups consists of 16-year olds (some of them are 17 years old) while the youngest group consists of 13 years old learners.

These findings raise several questions. Is this attributable to the learners' level of language proficiency? At what age do children acquire L1 figurative expressions? How does this affect children's ability to produce L2 semantically burdensome collocations? Starting with the difference in the semantic properties of the collocations produced by the four learner groups, I want to discuss this within the literature on the production of figurative expression focusing more on the role of the age of the learners. The aim is to explain the effect of learners' age and knowledge of L1 figurative language on the production of semantically burdensome collocations.

Various studies have found that the receptive and productive knowledge of figurative language correlates with age and years of schooling as well as being linked to other linguistic abilities (Bennelli et al, 2006; Vulchanova, Vulchanov & Stankova, 2011). According to Bennelli et al (2006), these other linguistic abilities include such thing as meta-linguistic

awareness and the ability to draw inference from context. Metalinguistic awareness is the "ability to reflect consciously on the nature and properties of language" (van Kleeck, 1982: 237). It is "the ability to focus on linguistic form and to switch focus between form and meaning" and it is "made up of a set of skills or abilities that the multilingual user develops owing to his/her prior linguistic and metacognitive knowledge" (Jessner, 2008: 275). This involves the understanding that language goes beyond the meaning, that words are separable from their referents and that language has a structure that can be manipulated (Mora, 2001). Learners would require this understanding to produce collocations with meanings beyond the meaning of each word in the collocational structure. The question then would be, what is the relationship between metalinguistic awareness and linguistic knowledge in second language learners and how much metalinguistic awareness do young learners have?

Alipour (2014: 2640) discovered a positive relationship between learners' metalinguistic knowledge and their "ability to correct, describe, and explain L2, and their proficiency in L2". This suggests an increase in learners' "metalinguistic awareness may increase the potential advantage of knowing two languages when learning a third" (Thomas, 1988: 235). If metalinguistic awareness involves the understanding that language goes beyond the meaning, that words are separable from their referents and that language has a structure that can be manipulated, it is plausible to draw a link between metalinguistic awareness and the production of semantically burdensome collocations. This is because semantically burdensome collocations are figurative. They have meanings that go beyond the literal meaning of the co-occurring lexical items. If metalinguistic awareness has some positive effect on L2 acquisition including the acquisition of figurative language, how is metalinguistic awareness developed as learners advance in age?

There is a strong evidence for consistent and applicable metalanguage awareness by age 7 or 8 (Saywitz & Cherry-Wilkinson, 1982). A study by Edwards and Kirkpatrick (1999) to determine if a developmental order exists in the metalinguistic ability of children to make judgments about the form of language while simultaneously attending to a meaningful linguistic context reveals a major shift in metalanguage ability occurring between 7 and 8 years of age. They discovered that children between the ages of 8 and twelve responded correctly to more items and at significantly faster rates than the children in ages 4 to 7. However, adults outperformed the children on all tasks, showing that metalanguage

development continues beyond childhood. What this suggests is that while children have metalinguistic ability at an early age, it is still in the process of developing. But what does this mean in terms of the effect of age on the acquisition of metaphor and how does that explain the reason why the younger and least proficient of my participant produced fewer semantically burdensome collocations – including metaphorical collocations?

Various studies have shown that L1 children acquire metaphoric language at very early age (Johnson & Pascual-Leone, 1989; Waggoner, Palermo & Kirsh, 1997; Wiśniewska-Kin, 2017) with children aged 11 to 12 able to reliably interpret most types of metaphors, even those that require fairly precise conceptualization (Billow,1975; Winner et al., 1976). According to Waggoner, Palermo and Kirsh (1997), children may interpret any combination of words metaphorically if a predictive enough context is present, regardless of the meanings of the words taken by themselves. This suggests they could interpret metaphoric collocations which are combination of words. What the above suggests is that L1 children have sufficient metalinguistic awareness to comprehend and produce metaphor. However, a study by Johnson and Pascual-Leone (1989) on developmental levels of processing in metaphor interpretation shows processing score increased with age in a predictable way. And "the ability to understand and produce metaphor in the L1 is related to the ability in the L2" (Littlemore, 2010: 302).

How could this be related to the production of semantically burdensome collocations? All the learners in my study are young learners who most likely have limited metalinguistic awareness and subsequently limited ability to produce L2 collocations with figurative meaning. As I have pointed out above, the knowledge of figurative expressions correlates with age, and it seems that the substantial gap in the semantically burdensome collocations produced by the learners and the native speakers may be the function of their language proficiency as well as their age. But the gap may equally be more of a function of their age rather than their L2 proficiency. The older learner group produced more collocations with figurative meaning and the number of such collocations recedes across the other three learner age groups. The link between age and metalinguistic awareness on the one hand and the likely link between metalinguistic awareness and the production of collocations with figurative meaning on the other hand seem to explain why the production of this types of collocations is non-existent in the texts produced by the youngest group of learners.

Macis and Schmitt (2017) investigate one hundred and seven, 18 – 36 years old Chilean Spanish-speaking English learners' knowledge of the figurative meanings of 30 collocations. Their result shows they have limited knowledge of idiomatic meaning of collocations, with a mean score of 33% correct. Generally, whether with younger learners or older learners, lexical items including collocations with idiomatic meaning are problematic for learners (Littlemore et al, 2011). While the fact that collocation is problematic for learners is well attested in the literature, we do not seem to know enough about the semantic properties of collocations produced by L2 learners. Most of the collocational studies in the literature have focused mainly on collocations with literal meaning.

9.4 Collocational Errors: A Window on L2 Mental Lexicon

This section further discusses the nature of the collocational errors and what they seem to reveal about their L2 mental lexicon. I will expand on the role of interlexical and intralexical factors in the production of collocations with a focus on clang associations and congruency. Clang associations, as I have said earlier on, are responses that have phonological resemblance to the stimulus words while polysemy means the capacity of a co-occurring word to have more than one meaning. I will attempt to discuss these within the literature on word association, L2 mental lexicon and relate them to Jiang's (2000) Model of Vocabulary Acquisition.

9.4.1 Clang Associations

One of the most frequent errors in the collocations produced by the learners in this study is clang associations – both phonological and orthographic clang. They are present in the collocational errors produced in three out of the four learner sub-corpora (NILECORP-A2, NILECORP-B2 and NILECORP-C1). The most proficiency group of learners which produced the highest numbers of well-formed collocations also produced the highest numbers of clang expressions. This was followed by the second most proficient groups and then the least proficient group. What we have here is the two most proficient groups (NILECORP-B2 and NILECORP-C1) producing the highest numbers of clang associations. Twelve out of the

144 instances of the non-teacher-norms verb noun collocation representing 8.3% of the unacceptable verb noun collocations produced by the most proficient group (NILECORP-C1) are clang associations. The second most proficient group (NILECORP-B2) have 49 instances of non-teacher-norms verb noun collocations out of which 30, representing 61.2% are clang associations. The least proficient group (NILECORP-A2) which produced the fewest well-formed collocations have seven instances of non-teacher-norms verb noun collocations. Two out of the seven unacceptable verb noun collocations representing 28.5% are clang associations. This means forty-four representing 20% of the 220 instances of non-teacher-norms verb noun collocations produced collectively the learners are responses that have phonological resemblance to the stimulus words.

In NILECORP-C1, there are thirteen instances where the learners are supposed to produce 'impart knowledge' but they produced *'impact knowledge'* ten times. They were also five instances in the corpus where it was appropriate to produce 'contract disease' but two of those instances, they instead produced *'contact disease'*. In NILECORP-B2, there are twenty-one instances where 'impart knowledge' is the appropriate collocation but in all those instances, they produced *'impact knowledge'*. There are also fifteen instances where the appropriate collocation is 'prescribe drug' but they produced *'describe drug'* seven times. Unlike the most proficient group, all the two instances where the collocation 'contract disease' is the appropriate form, they produced *'contact disease'*. The least proficient group (NILECORP-A2), produced 'wash television' twice instead of *'watch television'*. It is important to point out that the 'tʃ' sound as in /wɒtʃ/ (watch) does not exist in Yoruba language. For most Yoruba speakers, when they pronounce 'watch', they actually pronounce it as /wɒʃ/ (wash). This may be an additional layer of complications in the acquisition process for the learners. The production of so many clang expressions seems to be indicative of something in the learners' L2 mental lexicon.

The learners' responses to the stimulus words as could be seen above have been phonologically based rather than semantic. Besides, orthographically, the spelling of the words look so similar to the correct collocates. A clang, as in the case of 'wash' in 'wash television' above, have both orthographic and phonological resemblance to 'watch' but has no semantic connection to 'watch' which is the right collocate as in 'watch television'.

McCarthy's (1990: 41) explanation for L2 speakers' tendency to give clang responses is that the learners "may for a long time lack the ability to make instantaneous collocational associations, and may be more inclined to associate L2 words by sound similarities". This suggests that the organisation of their L2 mental lexicon at this stage is, to some extent, phonologically based which explains why they produce so many clang associations. It also indicates limited L2 semantic knowledge. Their limited L2 semantic knowledge results in the production of clang expressions which have no semantic relation to the appropriate collocate as in 'describe drug' for 'prescribe drug'. Though there are some similarities in the pronunciation of these words, there is no similarity in their meaning. They focus on the form of those words rather than their meaning. This seems to support various findings in word association literature which suggest that clang associations occurred more at early stage of L2 development (Meara, 1978, 1983; Namei, 2004). Various studies in word association have showed that unlike the L1 mental lexicon which is organised mainly on a semantic basis, the L2 mental lexicon is phonologically based in the early stage of development which is indicative of limited lexical knowledge (Meara, 1978, 1983; Namei, 2004; Zareva, 2007; Zhang & Nannan, 2014).

If L2 mental lexicon is form rather than meaning-based at the early stage of development, at what point of the developmental stage will it change to be more semantically based? The production of the collocation 'impart knowledge' across two proficiency groups may shed some light on this. There are 21 instances of that collocation in NILECORP-B2 and the learners produced 'impact knowledge' in all the 21 instances choosing a collocate that has phonological resemblance to the word *'impart'* but bears no semantic semblance to the word *'impart'*. The same collocation was produced 13 times in NILECORP-C1. Ten times the produced 'impact knowledge' and 'impart knowledge' three times. We can see a marginal shift from focus on form to focus on meaning as the learners' proficiency increases. What this tends to suggest is that the organisation of L2 mental lexicon shifts to be more semantically based later in the acquisition process. The sharp drop in the clang associations from 61.2% in NILECORO-B2 to 8.3% in NILECORP-C1 further suggests a shift in their lexical development as their proficiency increases. This is a shift from focus on form to focus on meaning in the acquisition process. The production of more clang expressions at lower level also seem to suggest a correlation between proficiency and clang production but one cannot make that conclusion on the basis of this study as clang associations are present in all but one

of the four learner groups. But what does this reveal about how lexical entries evolve in the learners' L2 mental lexicon?

By producing this many clang associations, it seems the learners, at early stages of their vocabulary acquisition process, focus on the formal features of the words. By producing words that have both orthographic and phonological resemblance (though seems to be based more on phonological resemblance than orthographic resemblance) to the stimulus words but no semantic semblance to the right words suggests that not much semantic information has been created and established in their mental lexicon. This seems to support Jiang's (2000) psycholinguistic model of vocabulary acquisition. He sees L2 lexical acquisition as consisting of three stages. His model postulates that at the initial stage – the formal stage, lexical entry with formal specifications are established. What the learners in this study seem to have done as the production of many clang associations suggest is focus on the formal specifications of the words. If they had focused on the semantic property of the words, they might not have produced these combinations.

Let us consider the production of 'impact knowledge' instead of 'impart knowledge' by the learners in NILECORP-B2 and NILECORP-C1. In all the 21 stances of the collocation 'impart knowledge' in NILECORP-B2, the learners produced the clang association 'impact knowledge'. But in NILECORP-C1, the same collocation was produced 13 times out of which it was produced correctly three times. What this suggest is that the learners in NILECORP-B2 focus on the formal features of the word 'impact' which has phonological resemblance to the word 'impart'. If the semantic information of this word had been registered in their L2 lexicon, they would have been able to differentiate the difference between the two words and they would not have produced the clang association. In NILECORP-C1 on the other hand, their production of the collocation correctly three times suggests a gradual progression from the formal stage toward integration stage where "semantic, syntactic, morphological as well as formal specifications about an L2 word are established within the lexical entry" (Jiang, 2000: 53). But will there be any time in the developmental process when L2 words that have phonological resemblance to the stimulus words in the production of collocation be less problematic? As a study by Pajak, Creel & Levy (2016: 1) show, "adults of particular L1 backgrounds have difficulty learning similar-

sounding L2 words that they can nevertheless discriminate perceptually". While learners at the initial stage of acquisition focus on the form of the word with none or little focus on their semantic specifications, words that sound similar pose additional challenge to L2 learners.

9.4.2 The role of Congruency and Frequency of Input in the Production of Collocations

The study has revealed from the pilot study to the main study that the singular most influential factor in the production of both well-formed and non-teacher-norms collocations is congruency. Most of the well-formed collocations produced by the learners are congruent and most of the non-teacher-norms collocations are incongruent. L1 negative transfer is the biggest source of L2 collocational errors across the four proficiency groups, and this mainly occurs when the structure is incongruent. Across all the four proficiency levels, congruency has been a facilitating factor while incongruency has been an inhibiting factor. Also, across all the proficiency groups, learners seem to produce correctly collocations that are frequently used in the Nigerian speech community than the ones that are less frequently used. While there are many complex agents that influence language acquisition, frequency of input and congruency appear to be the strongest influence in this young Yoruba-speaking Nigerian learners' production of collocations. I will now attempt to explain the influence of frequency on their production of collocations using the usage-based model of language acquisition (Tomasello, 2003).

Tomasello (2003: 69) summarises his usage-based approach to linguistic communication in the two aphorisms: "meaning is use [and] structure emerges from use". At the heart of the model, which has been primarily used in L1 studies, is the view of language acquisition as being mainly inductive and experience-driven process. What this suggests it that the frequency with which learners encounter language structures plays important role in the emergence of the language system. What this means for these learners is that through frequent use in the Nigerian speech community, certain collocational structures which are related to semantic and phonological or even orthographic structures (basically form-meaning mappings) become automatized – automatically retrievable by these learners of English. The model proposed that children come to the process of L1 language acquisition equipped with

two sets of cognitive skills namely: intention-reading which is the functional dimension and pattern-finding which is the grammatical dimension. This means children must "discern the goals or intentions of mature speakers when they use linguistic conventions to achieve social ends, and thereby to learn these conventions from them culturally" (ibid: 69 – 70). The second cognitive skill is what children need to do to enable them produce beyond the individual utterances they hear people use in their speech community to "create abstract linguistic schemas or constructions" (ibid: 70).

While not suggesting that the L2 learners are learning the language exactly the way L1 children would, as Ellis (2006a: 110) has cautioned that there are many factors that "filter and colour the perception of the second language", there is, however, a pattern in the findings that suggests some similarities. By producing more of the collocations that are frequently used in Nigeria – some of them which are peculiar to Nigeria and incongruent – the learners seem to have, in the course of frequent encounter with these structures, "discern the goals or intentions of mature speakers when they use linguistic conventions to achieve social ends, and thereby to learn these conventions from them culturally". And by having difficulty producing less frequently used collocations in Nigeria, could be because they are yet to have enough encounters with the collocational structures to do form-meaning mapping and ultimately have the structures entrenched in their mental lexicon. This seems consistent with Durrant and Schmitt's (2009: 157) findings "that non-native writers rely heavily on high-frequency collocations, but that they underuse less frequent, strongly associated collocations (items which are probably highly salient for native speakers)". Their findings also seem consistent with usage-based models of acquisition. Besides, could it be that young L2 learners behave like L1 children in their language development?

But then how does this model account for the fact that most proficient group of learners (NILECORP-C1) produced more collocational errors? Meanwhile, note that the same group produced more well-formed collocations and more incongruent collocations. 'Pattern-finding' which is the second cognitive skill the learners are equipped with in the model may account for this. It seems these learners, having done more 'pattern-finding' in the collocational structures they frequently hear in their speech community (probably more than the other three groups), were emboldened to "create abstract linguistic schemas or

constructions". In this process they produced more collocations – some of them well-formed and some of them not acceptable.

What this all means is that the acquisition of L2 collocations seems primarily based on the learners' exposure to the target structure in use and that they induce the 'rules' (collocations seem more arbitrary than rule-based) of their L2 from the patterns they are exposed to by employing cognitive mechanisms (Ellis & Wulff, 2014). This results in, to some extent, the production of well-formed collocations and sometime also results in the production of unacceptable collocations because the learners, most likely not fully awareness of the restriction on word combination, combine words that are not conventionally combinable.

Having said that, incongruency of the collocational structures add additional layer of difficulty to the learners' collocational development. The results across the four proficiency levels have shown that the learners have difficulty producing incongruent collocations. In this study, the production of incongruent collocations increases as their proficiency increases while their production of congruent collocations decreases as their proficiency increases. The least proficient group barely produced incongruent collocations. Jiang's vocabulary acquisition model, which is based on an extensive review of the existing literature, proposes a three-step process for L2 vocabulary acquisition. According to this model, the first step in vocabulary acquisition consists of creating an L2 entry that is linked to a corresponding L1 word, followed by a stage where learners integrate semantic, syntactic and morphological specification into the lexical entry appropriately morphologically and phonologically/orthographically but very much remains L1-like in respect to semantics and syntax. In Jiang's view, the third stage of vocabulary acquisition is achievable through more exposure to the L2 input which will result in gradual replacement of L1-based knowledge at the lemma level with more L2-based knowledge to create a lexical entry which is "very similar to a lexical entry in L1 in terms of both representation and processing" (Jiang, 2000: 53).

So how do the usage-based model and the vocabulary acquisition models account for this? The least proficient groups (NILECORP-A2 and NILECORP-B1) which produced the fewest incongruent collocations seem be at the stage where they map L2 entry (collocation

structures) into their existing lexical system which corresponds to the initial stage of Jiang's (2000: 51) model where "the use of L2 words involves the activation of the links between L2 words and their L1 translations". And because there seems to be either none or very weak link between L2 incongruent collocations and the learners L1 mental lexicon, they avoid the production of incongruent collocations. This stage seems to correspond to Tomasello's 'intention-reading' stage where the learners are connecting the language structures they hear around them to meaning (form-meaning mapping). The most proficient groups, on the other hand, are somewhere in between stage two and three of Jiang's model and seem to be moving in and out of Tomasello's 'intention-reading' and 'pattern-finding' stages. In the second stage of Jiang's model, as "experience in the L2 increases, strong associations are developed between L2 words and their L1 translations" which means "simultaneous activation of L2 word form and the lemma information (semantic and syntactic specifications) of L1 counterparts in L2 word use" (ibid: 51). While at his third stage, "the semantic, syntactic and morphological specifications of an L2 word are integrated from exposure and use and integrated into the lexical entry" (ibid: 53). This seems to correspond to the pattern-finding stage in the usage-based model. These entrenched specifications which are integrated from exposure and use and integrated into the learners' lexicon enable them to produce more incongruent collocations. But throughout the L2 collocations acquisition process, the learners will at various times have recourse to their L1 to produce incongruent collocations if they have never had enough exposure to such collocations. This probably explains why as various studies have shown, even advanced learners have difficulties producing incongruent and less frequent collocations (Nesselhauf, 2005; Laufer & Waldman, 2011; Durrant & Schmitt, 2009). I will now discuss the principal findings within the theoretical framework of the Revised Hierarchical Model of bilingual language processing.

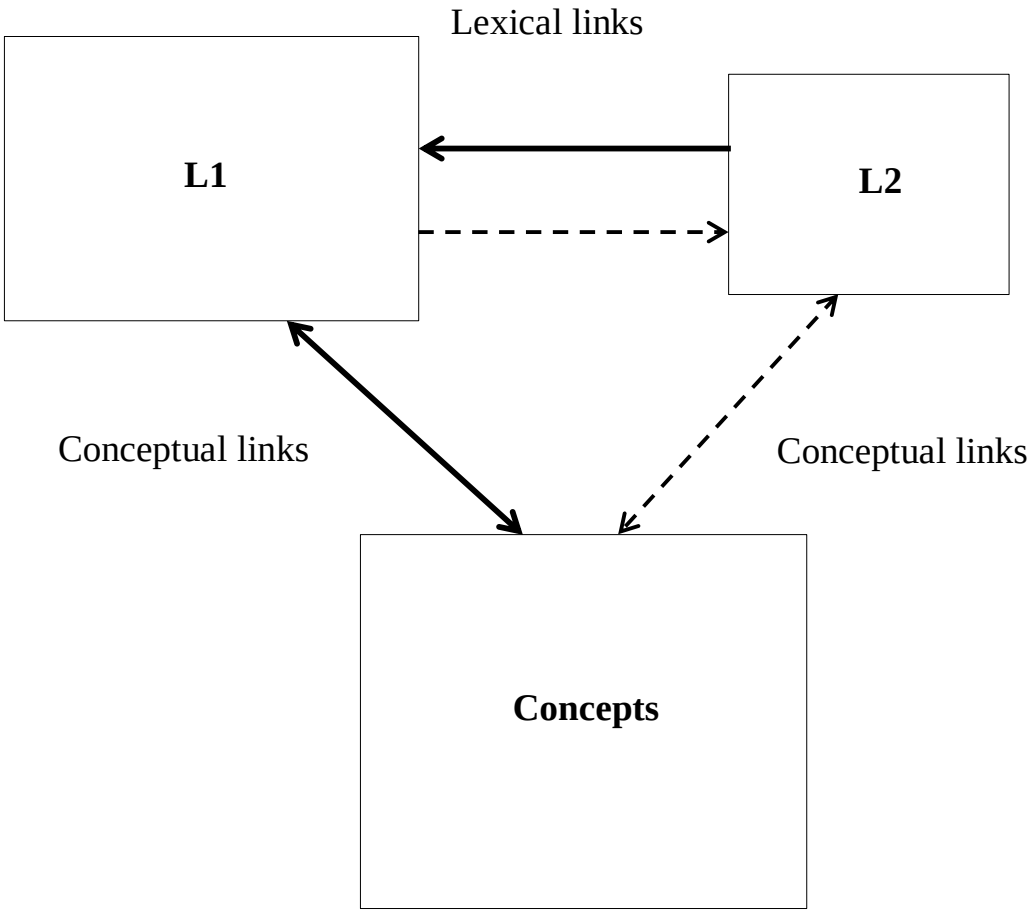
9.5 Production of L2 Collocations and the Revised Hierarchical Model

As I have said earlier, the often-cited Kroll and Stewart's (1994) Revised Hierarchical Model which "explains longer translation latencies from L1 to L2 (forward translation) than from L2 to L1 (backward translation) as an underlying asymmetry in the strength of the links between words and concepts in each of the bilingual's languages" (Kroll et al, 2010:373) assumes two levels of representations – lexical and conceptual. It accommodates independent lexical

representations for L1 and L2 with a shared conceptual representation. The model assumes links between L1 and L2 at the lexical level and a direct access from the form to the meaning in L1 and L2. According to this model, both the lexical and conceptual links are active in the bilingual memory. However, the strength of the links differs as a function of fluency in L2 and relative dominance of L1 over L2. The L1 is hypothesized to have privileged access to meaning because it is more developed and larger as the diagram below shows, while the L2 is thought to be more likely to require mediation through the L1 translation equivalent until the bilingual acquires sufficient skill in the L2 to access meaning directly. (Kroll et al, 2010). What this suggests is that as the L2 proficiency increases, the links between L2 words and concepts become stronger which means less use of L1 as a mediational tool for the production of L2 words, and learners begin to rely more on direct links – conceptual mediation.

In nutshell, as the diagram of the Revised Hierarchical Model below indicates, two routes lead from an L2 word form to its conceptual representation. One is the word association route, where concepts are accessed through the corresponding L1 word form (represented by the thick arrow from the L2 box to the L1 box and then to the concepts box), and the concept mediation route, with direct access from L2 to concepts (represented by the dotted arrow from the L2 box to the concepts box).

Revised Hierarchical Model of bilingual lexicon



As Pavlenko (2009: 143) rightly points out, "the unique strength of [this model] is in capturing the developmental change in linking between L2 and L1 word forms and lexical concepts". However, this model's assumption of a unified and stable conceptual store does not account for cases of partial and complete non-equivalence. My aim in this section is to discuss how the overall findings of this study (the production of congruent and incongruent collocations across the four proficiency groups, the nature of the collocational errors) play out in this model. I will also attempt to discuss, on the basis of this model, whether the collocational sequences are stored as a whole or as separate words.

The first set of principal findings of this study which are relevant to this model are that the learners across all proficiency levels produced more congruent collocations than incongruent collocations; that the production of collocations increased in tandem with proficiency

increase; and that the production of incongruent collocations increased as proficiency increased while the production congruent collocation decreased as proficiency increased. How can the Revised Hierarchical Model help to explain these findings? Various studies have suggested that the relationship of lexical/conceptual equivalence or near equivalence (cross-linguistic similarity) presents no difficulties for L2 vocabulary learning (Laufer & Eliasson, 1993; Pavlenko, 2008a; Pavlenko, 2009). This is because what L2 learners need to do is to link L2 word forms to already established lexical concepts as long as they subjectively perceive the concepts in question to be similar and this would result in positive L1 transfer facilitating the process. The production of seemingly disproportionately high numbers of congruent collocations in this study seems to suggest that in the learners' bilingual lexicons, the L1 is larger than the L2 as the model assumes, and the production of the collocations are largely forward translation (L1 to L2). If the two lexicons had been equal, the learners might have produced a proportionate number of both congruent and incongruent collocations.

Kroll and Stewart's (1994) model suggests that as the L2 proficiency increases, the links between L2 words and concepts become stronger which means learners begin to rely more on direct links – conceptual mediation and less use of L1 as a mediational tool for the production of L2 words. This developmental change in the link between L2 and L1 word forms and lexical concepts could explain why the production of collocations increased in tandem with proficiency increase. This could also explain why the production of incongruent collocations increased as proficiency increased while the production congruent collocation decreased as proficiency increased. The stronger link between L2 word forms and lexical concepts as proficiency increases means the most proficient groups have acquired sufficient skill in the L2 to access meaning directly and thereby rely less on L1 mediation in the production of L2 collocations. This seems to explain why the production collocations with no L1 equivalents increased as proficiency increased. However, this is not the case with the least proficient learners with a weaker link between L2 word forms and lexical concepts. It seems they have not acquired sufficient L2 skill to access meaning directly, they relied heavily on their L1 to produce the L2 collocations and therefore mainly produced collocations with L1 equivalents, with the A2 group production no incongruent collocations.

The second set of principal findings of this study which are relevant to this discussion within the Revised Hierarchical Model are that incongruency is the greatest cause of difficulty in L2 collocations production with all the unacceptable collocations produced by the learners being incongruent apart from one; and that L1 negative transfer is the biggest source of L2 collocational errors. The only congruent collocation which was produced wrongly was in the borderline between congruent and incongruent because of the polysemous nature of its collocate. This collocation is '*tell a story'* which some learners produced at *'talk a story'*. In Yoruba language, the lexical equivalent of the English verbs *tell, talk* and *say* is 'so'. This Yoruba verb 'so' is used in every context where any of the three English verbs (tell, talk and say) are used. This makes it harder for least proficient learner with weak link between L2 word forms and lexical concepts to select the right collocate for *story* as in *'tell a story'.* Meanwhile, the fact that L1 negative transfer is the biggest source of L2 collocational errors seems to be an evidence of a less established L2 lexicon compared to L1, and the learners seems to translate "from L2 to L1 (backward translation) as an underlying asymmetry in the strength of the links between words and concepts in each of the bilingual's languages" (Kroll et al, 2010:373). And as the model suggests, the locus of the asymmetry is at the lexical level. But the scale of the difficulties the groups of learners in this study have with incongruent collocations seems to question the assumption in Revised Hierarchical Model that bilinguals have a shared and stable conceptual store. This assumption does to accommodate cases of partial or complete non-equivalence which may either partially map on to the L1 partially (as in the case of '*shed blood'* which I discussed earlier) or not map on to the L1at all. There are some language-specific and culture-specific linguistic categories which are not shared in both languages (Yoruba and English) which means only one of the two languages may have the necessary word forms (Pavlenko, 2003). The implication of this is that the activation of the lexical links in one language would fail resulting in the production of unacceptable collocations.

Pavlenko's (2009) Modified Hierarchical Model which retains every aspect of the Revised Hierarchical Model but modifies the conceptual links to accommodate L1-specific categories, shared categories and L2-specific categories seems to better captures bilingual mental lexicon. Essentially, conceptual equivalence (shared categories) facilitates vocabulary learning, in this case, L2 collocations through positive transfer. Whenever L2 learners are

able to map form to meaning as in the case of conceptual equivalence linking L2 words and already existing concepts, they have little or no difficulty in producing the correct collocational structure. In the case of partial equivalence like 'shed blood' which I discussed earlier, the production of the L2 structure seems facilitated through partial overlap resulting in positive transfer through conceptual restructuring. However, in the case of conceptual non-equivalence where the linguistic category of the L2 does not have a counterpart in the L1, there is a greater possibility of producing unacceptable collocational structure.

The other thing that may come to mind is the question of whether the collocational sequences are stored as a whole or as separate words. Though, it is difficult to decide this on the basis of my data, various researchers have suggested they are stored and retrieved as a whole. Palmer's (1933, p. i) definition of collocation describes it as "a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts". This view was shared by Wray (2002: 9) who describes formulaic sequence as "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored, retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar". According to Ellis (1996: 111), it has long been acknowledged that a number of linguistic strings in our languages are treated like single "big words" which suggests they are regarded as "single choices, even though they might appear to be analysable into segments" (Sinclair, 1991: 110). My data which shows learners produced more of the collocations that are frequently used in their speech community than the less frequently used ones seem to suggest they stored the collocation as a whole as a result of frequent co-occurrence and retrieved them as a whole from their memory whenever needed.

Chapter Ten

Conclusion

10.0 Introduction

This study was born out of the need to fill the gap in the literature on L2 collocations within the concept of World Englishes; and Learner Corpus Research in Nigeria. The first aim was to build a half a million words learner corpus of Nigerian English. This would be a precursor to a bigger (open access) 10 million words multilevel learner corpus representing various Nigerian L1 speakers designed in a way that will allow for comparative study of various L1 learners of Nigerian English. The second aim was to gain a better understanding of the collocational competence and development of learners of English in a context where a nativized variety of English is the second language – context that can be likened to the learning of a language through immersion. But above all, investigating collocational competence and analysing collocational errors not based on exonormative models but on endonormative model – reflecting the sociolinguistic reality of the English language use in the Nigerian speech community. Studies on L2 collocations competence and development in the existing literature have not investigated the existence of collocations in the emerging varieties of English which may not exist in the prestigious varieties. For the so-called native speakers, such collocational combinations may be infelicitous, but they are variety makers of the new Englishes.

One of the greatest achievements of this study is the building of the 516, 917 words multilevel Nigerian Learner Corpus of English NILECORP and the assignment of proficiency levels to the corpus texts. The assignment of proficiency levels to the corpus data on the basis of the Common European Framework of Reference for Languages (CEFR) allows for researchers who are not familiar with the Nigerian context to be able to make sense of the findings of this study as well as to replicate this study in another context. Furthermore, this pioneering learner corpus will be used for various linguistic enquiries beyond this thesis which will result in the publications of several peer-reviewed articles. Besides, the

experience of compiling the corpus has equipped me with the necessary skills to build the 10 million words Learner Corpus of Nigerian English. Apart from the learner corpus, this study has successfully, for the first time, conducted a comprehensive investigation of the collocational competence and development of Nigerian learners of English. The successful assignment of proficiency levels to the corpus texts means this study was able to examine the development of collocational knowledge across four difference proficiency levels – something that has never been done in Nigeria. Because the texts of the corpus are written by Yoruba speaking participants, I was able to determine which collocations are congruent and which ones are not. This also made it possible to account for the sources of the collocational errors. By using frequency data from the Nigerian component of GloWbE, it was possible to investigate the effect of frequency of certain collocational structure in the local context on the learners' production of collocations. While most studies have ignored the semantic properties of collocations, this study successfully investigated this aspect of collocations by comparing collocations produced by the L2 learners and native speakers, and by investigating this across proficiency levels. All these have produced findings that were not known until now about collocations in World Englishes, and the collocational competence and development of Nigerian learners of English. In a nutshell, this study has contributed to our understanding of collocation as a linguistic concept, particularly the acquisition and usage of collocations within the context of World Englishes. The next section provides a summary of the findings of this study.

10.1 Summary of Findings

The second aim of this study has been to investigate the production and use of collocations by Nigerian English learners. The investigation started with a comparative analysis of the collocations produced by the most proficient of the four groups of learners and the native speakers. The first finding was that, in quantitative terms, relatively advanced learners of English from an English as a second language context where the learners have frequent exposure to the input outside the classroom can produce as many collocations in a written text as native speakers do. The learners produced more congruent collocations (63.1%) than incongruent collocations (36.9%). The second comparative analysis focused on the linguistic complexity – the collocational span and the structural properties of the constituents of the verb noun collocations produced by the native speakers and the L2 learners. It was

discovered that while the learners produced almost as many collocations in the corpus as the native speakers did, in terms of the length of the collocational structures (the collocational span), the ones produced by the native speakers are noticeably different from the ones produced by the learners. The native speakers overwhelmingly produced more long span collocations than the L2 learners. The analysis also revealed a wide gap in the structural complexity of the constituents of the verb noun collocations produced by native speakers and the learners, and that the learners tend to overuse a few favourite structurally complex verb noun collocations.

A comparison of the production of collocations with additional nuances and associations by both groups showed that 8.2% of all the collocations produced by the L2 learners are semantically burdensome while 9.7% of the collocations produced by the native speakers are semantically burdensome. However, if the semantically burdensome collocations produced by the learners were to be put in a single continuum within the same processing system from fully transparent to fully opaque, they would be on the lower end of opacity while the ones produced by the native speakers would be on the upper end of opacity. This simply means there is a gap between the collocations produced by the learners and the native speakers in terms of using collocations to reflect various shades of meaning from fully transparent to fully opaque. In a nutshell, the difference between the collocations produced by the learners and the native speakers did not lie in the quantity but in the linguistic complexity – structural and semantic properties of the collocations.

The second research questions investigated the effect of frequency and exposure to input the learners' speech community affect the collocational production of the most proficient of the four learner groups. The analysis revealed that 81.2% of the incongruent collocational structures and 78.4% of the congruent collocational structures extracted from the learner corpus are frequently used in the Nigerian component of GloWbE. Overall, 80.2% of all the collocational structures produced by the learners are frequently used in the Nigerian component of GloWbE. But 19.8% of all the collocational structures produced that are not frequently used in the Nigerian component of GloWbE. It was concluded that: (1) frequency and exposure to input facilitate the productive knowledge of collocations, (2) frequency

trumps incongruency (3) but the production of collocation is not entirely the function of frequent exposure to input.

The third research question enquired into the relationship between language proficiency and the production of collocations comparing data across four proficiency levels. It attempted to find out if L2 learner's knowledge of collocations increases in tandem with their general proficiency in the English language. It examined the effect of proficiency on the production of: (1) congruent and incongruent collocations, (2) linguistically complex verb noun collocations and (3) collocations with additional nuances and associations. The findings point to a strong link between proficiency and the production of collocations. Production of collocations increased in tandem with proficiency increase. One of the most interesting findings is the reversal of relationship between the production of incongruent collocations and proficiency on the one hand and the production of congruent collocations and proficiency on the other hand. The production of incongruent collocations increased as proficiency increased while the production congruent collocation decreased as proficiency increased. This seems to indicate that L2 learners rely heavily on their L1 to produce L2 collocations, and this reliance thins out as they become more proficient. This accounts for their production of fewer incongruent collocations at the least proficient levels and increases as they become more proficient.

The second part of the analysis which focused on the linguistic complexity (the collocational span, the structural and semantic properties) of the collocation produced across the four learner groups revealed that three of the four proficiency groups consistently show an increase in the number of long span collocations as their proficiency increases. The two most advanced groups produced remarkably more structurally complex collocations than the two least proficient groups. There is also a link between proficiency and the production of collocations with figurative meaning. While the two most proficient groups produced a substantial number of collocations with figurative meanings, such collocations are almost non-existent in the text produced by the least proficient groups – B1 and A2.

The last main research question attempted to analyse all the unacceptable collocations produced by the learners with the aim of identifying, classifying and accounting for the errors. The analysis revealed that learners across the four proficiency levels have difficulty producing incongruent collocations. The most proficient group which produced more acceptable collocations than the others also produced the highest numbers of unacceptable collocations. This was considered a positive developmental process as it means the learners buoyed by their increasing proficiency were willing to take risk in their output resulting in the production of more collocations – many of them acceptable and some unacceptable. On the contrary, the least proficient learners stayed in their comfort zone which means fewer collocations were produced and fewer collocational errors were made. Finally, L1 negative transfer was the main source of collocational errors which suggests L2 learners regardless of their proficiency, recourse to their L1 to produce collocations particularly when the target structure is incongruent.

10.2 Limitation of the Study

I should stress that my study was limited to verb noun and adjective noun collocations. Initially, I wanted to consider Verb + Noun, Adjective + Noun, Adverb + Adjective, Verb + Adverb, Noun + Verb, and Noun + Noun but had to limit to two sub-sets because of the sheer volume of the collocations in the corpus. This, in essence, is not a limitation but an opportunity for further studies.

I should also make clear that the nature of my data does not allow me to determine whether some of the collocational errors were made by most of the participants or by a few individual learners. The assignment of proficiency levels to the corpus texts produced by group of learners as opposed to the texts produced by individual learners means the data does not account for possible individual differences in terms of language abilities.

10.3 Recommendations

One of the most striking findings of this study is that, in quantitative terms, relatively advanced learners of English (equivalent to CEFR C1) from an English as a second language context where the learners have frequent exposure to the input outside the classroom, can produce as many collocations in a written text as native speakers do. It is recommended that this study be replicated in another context to determine whether the ability to produce so many collocations is a function of the Nigerian context or something else.

Future research into L2 collocations might focus on World Englishes in other contexts and use endonormative model instead of exonormative model when decided which collocation are acceptable and which are not so as to account for localised collocations. Collocational studies in Nigeria could focus on other L1 speakers. A longitudinal study focusing on certain learners could shed more light on the development of collocational knowledge. It will also be interesting to investigate the use of collocations by native and non-native speakers across time periods.

References

Achebe, C. (1975). *Morning yet on creation day: Essays*. London: Heinemann.

Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography. *ITL-International Journal of Applied Linguistics*, *53*(1), 53-61.

Ajani, T. (2007). Is There Indeed a Nigerian English? *Journal of Humanities and Social Sciences 1 (1)*, Available at: http://www.scientificjournals.org /journals2007/ articles/1084.htm [Accessed 17 May 2014].

Akande, A. T., Adedeji, E. O., & Okanlawon, B. O. (2006). Lexical errors in the English of technical college students in Osun State of Nigeria. *Nordic Journal of African Studies*, *15*(1).

Alipour, S. (2014). Metalinguistic and linguistic knowledge in foreign language learners. *Theory and Practice in Language Studies*, *4*(12), 2640-2645.

Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, *4*(1), 71-83.

Al-Zahrani, M. S. (1998). Knowledge of English lexical collocations among male Saudi college students majoring in English at a Saudi university. Unpublished.

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research*, *30*(2), 141-161.

Anyachonkeya, N., & Anyachonkeya, C. (2015). Features of Nigerian English. *The Melting Pot*, *1*(1).

Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and linguistic computing*, *7*(1), 1-16.

Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *Tesol Quarterly*, 535-556.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Bahns, J. (1993). Lexical Collocations: a contrastive view. *ELT Journal 47/1: 56 –63.*

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations?. *System*, *21*(1), 101-114.

Bambose, A. (1982). “Standard Nigerian English: Issues of Identification” in Braj, B. Kachru (ed) The other Tongue; English Across Culture. Urbana: University of Illinois Press.

Bamgbose, A. (1995). English in the Nigerian environment. In Bamgbose, A., Banjo, A., & Thomas, A. (Eds), New Englishes (pp. 9-26). Ibadan: Mosuro Publishers.

Bamgbose, A., Banjo, A., & Thomas, A. (1995). *New Englishes. A West African Perspective.* Ibadan: Mosuro & The British Council.

Barfield, A. (2009). Exploring productive L2 collocation knowledge. *Lexical processing in language learners: Papers and perspectives in honour of Paul Meara,* 95-110.

Barfield, A., & Gyllstad, H. (2009). *Researching collocations in another language: Multiple interpretations*. Basingstoke: Palgrave McMillan.

Barlow, M., & Kemmer, S. (Eds.). (2000). Usage based models of language. Stanford, CA: CSLI Publications.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (Vol. 3, No. 1). New York: Wiley.

Bartsch, S. (2004). *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Gunter Narr Verlag.

Bartsch, S., & Evert, S. (2014). Towards a Firthian notion of collocation. Network Strategies, Access Structures and Automatic Extraction of Lexicographical Information. 2nd Work Report of the Academic Network Internet Lexicography, OPAL–Online publizierte Arbeiten zur Linguistik. Institut für Deutsche Sprache, Mannheim, to appear.

Begagić, M. (2015). English language students’ productive and receptive knowledge of collocations. *Explorations in English Language and Linguistics*, *2*(1), 46-67.

Benelli, B., Belacchi, C., Gini, G., & Lucangeli, D. (2006). ‘To define means to say what you know about things’: the development of definitional skills as metalinguistic acquisition. *Journal of Child Language, 33*(1), 71-97.

Benson, M. (1985). Collocations and idioms. *Dictionaries, lexicography and language learning*, 61-68.

Benson, M. (1990). Collocations and general-purpose dictionaries. *International Journal of Lexicography*, *3*(1), 23-34.

Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. John Benjamins Pub. Co..

Benson, M., Benson, E., & Ilson, R. F. (1986). *Lexicographic description of English* (Vol. 14). John Benjamins Publishing.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In Sinclair, J. M. (Ed.). *How to use corpora in language teaching* (Vol. 12). John Benjamins Publishing.

Berry-Rogghe, G. (1973). The computation of collocations and their relevance in lexical studies. *The computer and literary studies*, 103-112.

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, *8*(4), 243-257.

Billow, R. (1975). A cognitive developmental study of metaphor comprehension. *Developmental Psychology*, 11, 415-423.

Biskup, D. (1990). Some remarks on combinability: Lexical collocations. *Foreign language acquisition papers*, 31-44.

Biskup, D. (1992). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. In *Vocabulary and applied linguistics* (pp. 85-93). London: Palgrave Macmillan.

Bisson, M. J., Heuven, W. J., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language learning*, *64*(4), 855-877.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *Vigo International Journal of Applied Linguistics*, (11).

Boonyasaquan, S. (2009). An analysis of collocational violations in translation. ***วารสารมนุษยศาสตร์ ปริทรรศน์*** *(MANUTSAT PARITAT: Journal of Humanities)*, *27*(2).

Boscolo, P. (1991). Contexts for writing, writing in context. *European Journal of Psychology of Education*, *6*(2), 167-174.

Boulton, A. (2017). Data-Driven Learning and Language Pedagogy. *Language, Education and Technology*, 1-12.

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, *20*(2), 139-173.

Brown, D. F. (1974). Advanced vocabulary teaching: The problem of collocation. *RELC journal*, *5*(2), 1-11.

Brown, H. D. (2000). Principles of language learning and teaching (4th Ed.). White Plains, NY: Longman.

Burgin, M. (2016). Theory of Knowledge: Structures and Processes. Singapore: World Scientific Publishing Company.

Burnard, L. (2007). Reference Guide for the British National Corpus (XML Edition). Published for the British National Corpus Consortium by the Research Technologies Service at Oxford University Computing Services.

Campbell, G. L. (1991). *Compendium of the World's Languages* (Vol. 21991). London: Routledge.

Carlsen, C. (2009). 'Proficiency levels in learner corpora – a source of error or an asset in SLA-research'. Paper presented at GURT 2009,' Georgetown University Round Table, WA, USA, 13–15 March 2009.

Carlsen, C. (2012). Proficiency level—A fuzzy variable in computer learner corpora. *Applied Linguistics*, *33*(2), 161-183.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: a comprehensive guide; spoken and written English grammar and usage*. Cambridge University Press.

Chen, W. (2017). Profiling Collocations in EFL Writing of Chinese Tertiary Learners. *RELC Journal*, 0033688217716507.

Chen, Y. (2004). A corpus-based investigation of collocational errors in EFL Taiwanese high school students' compositions. *Theses Digitization Project*. 2579.

Choueka, Y. (1988). Looking for needles in a haystack or locating interesting collocational expressions in large textual databases. In *RIAO 88:(Recherche d'Information Assistée par Ordinateur). Conference* (pp. 609-623).

Christiana-Oluremi, A. O. (2013). Lexico-grammatical features of Nigerian English. *International Journal of English Language Education*, *1*(1), 261-271.

Clear, J. (1993). From Firth Principles: Computational Tools for the Study of Collocation. In M. Baker et al. (eds) Text and Technology. Amsterdam: Benjamins, 271-292.

Cohen, L., Manion, L., & Morrison K. (2000). *Research Methods in Education* (5th Edition). London: Routledge Falmer.

Collentine, J., & Freed, B. F. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in second language acquisition*, *26*(2), 153-171.

Coseriu, E. (1967). Lexikalische solidaritäten. Poetica, 1, 293-303.

Council of Europe. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.

Cowie, A. P. (1978). The place of illustrative material and collocations in the design of a learner's dictionary. *In honour of AS Hornby*, *127139*.

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied linguistics*, *2*(3), 223-235.

Cowie, A. P. (1992). Multiword lexical units and communicative language teaching. In *Vocabulary and applied linguistics* (pp. 1-12). Palgrave Macmillan, London.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.

Crystal, D. (2003). *English as a global language*. 2nd edition. Cambridge University Press.

Davies, M. (2013) *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE)*. Available online at https://corpus.byu.edu/glowbe/. [Accessed 10 January 2015].

Davies, M., & Fuchs, R. (2015). Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide*, *36*(1), 1-28.

Davoudi, M., & Behshad, A. (2015). Collocational use: a contrastive analysis of strategies used by Iranian EFL learners. *Theory and Practice in Language Studies*, *5*(12), 2646-2652.

De Klerk, V. (2002). Towards a corpus of black South African English. *Southern African Linguistics and Applied Language Studies*, *20*(1-2), 25-35.

Demir, C. (2017). Lexical collocations in English: A comparative study of native and non-native scholars of English. *Journal of Language and Linguistic Studies*, *13*(1), 75-87.

Doiz, A. & Elizari, C. (2013). Metaphoric competence and the acquisition of figurative vocabulary in foreign language learning. *ELIA*, (13), 47.

Dürmüller, U. (2008). Towards a new English as a foreign language curriculum for Continental Europe. In Miriam A. Locher & Jürg Strässler (eds.), *Standards and norms in the English language (pp. 239–253)*. Berlin and New York: Mouton de Gruyter.

Durrant, P. L. (2008). *High frequency collocations and second language learning* (Doctoral dissertation, University of Nottingham).

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*(2), 125-155.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations?. *IRAL-International Review of Applied Linguistics in Language Teaching*, *47*(2), 157-177.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163-188.

Ebrahimi-Bazzaz, F., Samad, A. A., bin Ismail, I. A., & Noordin, N. (2014). Verb-noun collocation proficiency and academic years. *International Journal of Applied Linguistics and English Literature*, *3*(1), 152-162.

Ecke, P., & Hall, C. J. (2014). The Parasitic Model of L2 and L3 vocabulary acquisition: evidence from naturalistic and experimental studies. *Fórum Linguístico*, *11*(3), 360-372.

Edem, E. D. (2016). A survey of the Syntactic Features of Educated Nigerian English. *International Journal of Academia, (2)1, 1 – 18.*

Edwards, H. T., & Kirkpatrick, A. G. (1999). Metalinguistic awareness in children: A developmental progression. *Journal of psycholinguistic research*, *28*(4), 313-329.

Ehineni, T. O. (2014). Lexical, structural and rhetorical features of Nigerian English print media. *Journal of Arts and Humanities*, *3*(11), 26-32.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, *24*(2), 143-188.

Ellis, N. C., & Wulff, S. (2014). Usage–based approaches to SLA1. *Theories in second language acquisition: An introduction*, *1*, 75.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages (pp. 79 – 114).* San Diego: Academic Press.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Enyi, A. U. (2015). Style of Nigerian English Conversation: A Discourse-Stylistic Analysis of a Natural Conversation. *European Journal of English Language and Literature Studies*, *3*(4), 42-53.

Evert, S. (2004). The Statistics of Word Co-occurrences: Word Pairs and Collocations.

Eyckmans, J. (2009). Towards an assessment of learners' receptive and productive syntagmatic knowledge. In A. Barfield & H. Gyllstad (Eds.), Researching collocations in another language: Multiple interpretations (pp. 139-152). Basingstoke: Palgrave Macmillan.

Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics, 33(4), 315-31.*

Fernández, B. G., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? *ITL-International Journal of Applied Linguistics*, *166*(1), 94-126.

Firth, J. R. (1951). General linguistics and descriptive grammar. *Transactions of the Philological Society*, *50*(1), 69-87.

Firth, J. R. (1957). Modes of meaning. In: Papers in Linguistics, 1934-1951. Oxford: Oxford University Press.

Firth, J. R. (1961). *Papers in Linguistics 1934-1951:* Oxford: Oxford University Press.

Fishman, J. A. (1992). Sociology of English as an additional language. *The other tongue: English across cultures*, *2*, 19-26.

Fontenelle, T. (1992). Collocation acquisition from a corpus or from a dictionary: a comparison. In *Proceedings I-II. Papers submitted to the 5th EURALEX International Congress on Lexicography in Tampere* (pp. 221-228).

Friday-Òtún, J. Ọ., & Ọmọ́léwu, C. Ọ. (2016). Collocative Syntagms in the Yorubá Language Usage. *Language*, *4*(2), 8.

Garside, R., Leech, G. N., & McEnery, T. (Eds.). (1997). *Corpus annotation: linguistic information from computer text corpora*. London: Routledge Taylor & Francis.

Gass, S., & Selinker, L. (2001). Second language *acquisition: An introductory course.* Mahwah, NJ: Lawrence Erlbaum.

Gelbukh, A., & Kolesnikova, O. (2013). Linguistic Interpretation. In *Semantic Analysis of Verbal Collocations with Lexical Functions* (pp. 85-92). Springer, Berlin, Heidelberg.

Gilquin, G., De Cock, S., & Granger, S. (2010). The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM.

Glaznieks, A., Nicolas, L., Stemle, E., Abel, A., & Lyding, V. (2014). Establishing a standardised procedure for building learner corpora. *Apples: journal of applied language studies*.

Goh, G. Y. (2011). Choosing a reference corpus for keyword calculation. *Linguistic Research*, *28*(1), 239-256.

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? *ITL-International Journal of Applied Linguistics*, *166*(1), 94-126.

Granger S. (2008). Learner Corpora in Foreign Language Education. In Van Deusen-Scholl N. and Hornberger N.H. (ed.) *Encyclopedia of Language and Education. Volume 4. Second and Foreign Language Education.* Springer, 337-351.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. *Phraseology: Theory, analysis, and applications*, *145*, 160.

Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. In Granger, S. (ed.) (1998). *Learner English on Computer*. Addison Wesley Longman: London & New York, 3-18

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, and S. Petch-Tyson (eds), *Computer learner corpora, second language acquisition and foreign language teaching*, Amsterdam: John Benjamins, pp. 3-33.

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. (2004). Computer learner corpus research: current status and future prospects. *Language and Computers*, *52*, 123-146.

Granger, S. (2012). How to use foreign and second language learner corpora. In A. Mackey & S. Gass (eds.) *Research Methods in Second Language Acquisition: A Practical Guide.* Malden: Blackwell, 7-29.

Granger, S. (2015). Contrastive Interlanguage Analysis: A reappraisal. *International Journal of Learner Corpus Research*, *1*(1), 7-24.

Granger, S. (Ed.). (1998). Learner English on computer. London: Addison Wesley Longman

Granger, S., Dagneaux, E. Meunier, F., & Paquot, M. (2009). International corpus of learner English. (2nd version). Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2013). Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead: Proceedings of the First Learner Corpus Research Conference (LCR 2011) (Vol. 1). Presses Universitaires de Louvain.

Groom, N. (2009). Effects of second language immersion on second language collocational development. In A. Barfield & H. Gyllstad (Eds.), Researching collocations in another language: Multiple interpretations (pp. 21-33). Basingstoke: Palgrave Macmillan.

Guo, X. (2006). *Verbs in the written English of Chinese learners: A corpus-based comparison between non-native speakers and native speakers* (Doctoral dissertation, University of Birmingham).

Gyllstad, H. (2007). Testing English collocations: Developing receptive tests for use with advanced Swedish learners. Lund: Lund University.

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter?. *Language Learning*, *66*(2), 296-323.

Ha, M. J. (2013). Corpus-based Analysis of Collocational Errors. *International Journal of Digital Content Technology and its Applications*, *7*(11), 100.

Hall, C. J. (1992). Making the Right Connections: Vocabulary Learning and the Mental Lexicon. Unpublished manuscript, Universidad de las AmÈricas, Puebla: ERIC Document Reproduction Service No. ED 363 128.

Hall, C. J., & Ecke, P. (2003). Parasitism as a default mechanism in L3 vocabulary acquisition. In *The multilingual lexicon* (pp. 71-85). Springer, Dordrecht.

Halliday, M. A. K. (1978). Language as Social Semiotic: The Social Interpretation of Language and Meaning. London: Edward Arnold.

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. English Language Series, London: Longman.

Halliday, Michael A. K. (1966). Lexis as linguistic level. In Charles E. Bazell, John C. Catford, Michael A. K. Halliday & R. H. Robbins (Eds.), In Memory of F. R. Firth, pp. 148–162. Harlow, U.K.: Longman.

Hama, H. Q. (2010). Major sources of collocational errors made by EFL learners at Koya University. *Unpublished Master's). Bilkent University, Ankara, Turkey*.

Hausmann, F. J. (1985). Kollokationen im deutschen wörterbuch. ein beitrag zur theorie des lexikographischen beispiels. *Lexikographie und Grammatik. Niemeyer, Turgen, Germany*.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development–a progress report. *C. Bardel, C. Lindqvist, & B. Laufer (Eds.)* L2 vocabulary acquisition, knowledge and use, 29-56.

Herbst, T. (1996). What are collocations: sandy beaches or false teeth? *English Studies*, 77(4): 379–93.

Heuboeck, A., Holmes, J., & Nesi, H. (2007). *The BAWE corpus manual*. Technical report, Universities of Warwick, Coventry and Reading.

Hoey, M. (2000). A world beyond collocation: New perspectives on vocabulary teaching. In M. Lewis (Ed.) *Teaching collocations (*pp. 224-245). Hove: Language Teaching Publications.

Holtz, M. (2007). Corpus-based analysis of verb/noun collocations in interdisciplinary registers. Proceedings of the Corpus Linguistics conference CL 2007. http://ucrel.lancs.ac.uk/publications/CL2007/paper/14_Paper.pdf [Accessed 10 October 2018].

Hornby, A. S. (1942). *Idiomatic and syntactic English dictionary*. Tokyo: Institute for Research in Language Teaching.

Hornby, A. S., Gatenby, E. V., & Wakefield, H. (1948). *A Learner's dictionary of current English*. London: Oxford Univ. Press.

Hornby, A. S., Gatenby, E. v. & Wakefield, H. (1963). *The advanced learner's dictionary of current English* (Vol. 965) London: Oxford University Press.

Hosseini, B., & Akbarian, I. (2007). Language proficiency and collocational competence. *The Journal of Asia TEFL*, *4*(4), 35-58.

Houston, J. E. (2001). *Thesaurus of ERIC descriptors*. Greenwood Publishing Group.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics*, *19*(1), 24-44.

Howarth, P. A. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making* (Vol. 75). Walter de Gruyter.

Hsu, J. Y., & Chiu, C. Y. (2008). Lexical collocations and their relation to speaking proficiency of college EFL learners in Taiwan. *Asian EFL Journal*, *10*(1), 181-204.

Hulstijn, J. H. (2007). The Shaky Ground Beneath the CEFR: Quantitative and Qualitative Dimensions of Language Proficiency. *The Modern Language Journal*, *91*(4), 663-667.

Hulstijn, J. H., Alderson, J. C., & Schoonen, R. (2010). Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them. *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, 11-20.

Hunston, S. (2006). Phraseology and system: A contribution to the debate. *Equinox Publishing*, 55-80.

Ifecheobi, J. N. (2016). Language use and the Mass Media: A Focus on Selected Nigerian Newspapers. *Awka Journal of English Language and Literary Studies*, *3*(1), 51-63.

Igboanusi, H. (2002). *A dictionary of Nigerian English usage*. Enicrownfit Pub.

Heid, U. (1994). On ways words work together – research topics in lexical combinatorics. In Proceedings of the 6th Euralex International Congress on Lexicography (EURALEX '94), pages 226–257, Amsterdam, The Netherlands

Ishikawa, S. (2013). ICNALE: the international corpus network of Asian learners of English. Available at: http://language.sakura.ne.jp/icnale/ [Accessed 10 November 2018].

Israel, P. C. (2014). Effects of Lexico-syntactic Errors on Teaching Materials: A Study of Textbooks Written by Nigerians. *International Journal of Education and Literacy Studies*, *2*(1), 75-81.

Jarvis, S. (2009). Lexical transfer. *The bilingual mental lexicon: Interdisciplinary approaches*, 99-124.

Jessner, U. (2008). A DST model of multilingualism and the role of metalinguistic awareness. *The modern language journal*, *92*(2), 270-283.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied linguistics*, *21*(1), 47-77.

Jindapitak, N. & Teo, A. (2013). The emergence of World Englishes: Implications for English Language teaching. *Asian Journal of Social Sciences and Humanities,* 2 (2), pp. 190-199.

John R. Firth, J. R. (1968). A synopsis of linguistic theory, 1930–55. In F.R. Palmer (Ed), Selected papers of J. R. Firth, 1952–1959 (pp. 168–205). Bloomington: Indiana University Press.

Johns, T., & King, P. (1991). Classroom Concordancing: English Language Research Journal, 4. *University of Birmingham: Centre for English Language Studies*.

Johnson, J., & Pascual-Leone, J. (1989). Developmental levels of processing in metaphor interpretation. *Journal of Experimental Child Psychology*, *48*(1), 1-31.

Jowitt, D. (1991). *Nigerian English Usage: An Introduction.* Ikeja: Longman.

Kaan, A. T., Amase, E. L. P., & Tsavmbu, A. A. (2013). Nigerian English: Identifying Semantic Features as Variety Markers. *IOSR Journal of Humanities and Social Science, 16*(5), 76-80.

Kachru, B. B. (1992). World Englishes: Approaches, issues and resources. *Language teaching*, *25*(1), 1-14.

Kachru, B. B. (Ed.). (1992). *The other tongue: English across cultures*. University of Illinois Press.

Kahane, S., & Polguere, A. (2001). Formal foundation of lexical functions. In *Proceedings of ACL/EACL 2001 Workshop on Collocation* (pp. 8-15).

Källkvist, M. (1998). Lexical infelicity in English: the case of nouns and verbs. *Perspectives on lexical acquisition in a second language*.

Kamshilova, O. N. (2017). Overuse In Learner Language: Frequency And Accuracy. *Russian Linguistic Bulletin*, 3(11), 28-31.

Kaufmann, S. (1999). Cohesion and collocation: Using context vectors in text segmentation. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 591-595). Association for Computational Linguistics.

Kennedy, G. (2014). *An introduction to corpus linguistics*. London: Routledge.

Kim, S. H., & Kim, J. H. (2012). Frequency Effects in L2 Multiword Unit Processing: Evidence From Self-Paced Reading. *TESOL Quarterly*, *46*(4), 831-841.

Kjellmer, G. (1987). Aspects of English Collocations in Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora. *Costerus*, *59*, 133-140.

Kjellmer, Göran (1990). A mint of phrases. In *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, Karin Aijmer and Bengt Altenberg (eds.), 111–127. London: Longman.

Kolesnikova, O., & Gelbukh, A. (2015). Measuring non-compositionality of verb-noun collocations using lexical functions and wordnet hypernyms. In *Mexican International Conference on Artificial Intelligence* (pp. 3-25). Springer, Cham.

Kperogi, F. A. (2007). Divided by a common language: Comparing Nigerian, American and British English. Retrieved from http://www.farooqkperogi.com/2007/09/divided-by-common-language-comparing.html [Accessed 20 may 2016].

Krishnamurthy, R. (2000). Collocation: from silly ass to lexical sets. In C. Heffer, H. Sauntson, and G. Fox (Eds.) Words in Context: A Tribute to John Sinclair on his Retirement (pp. 31 -47). Birmingham: University of Birmingham.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149-174.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149-174.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373-381.

Kumar, V., Stubbs, A., Shaw, S., & Uzuner, Ö. (2015). Creation of a new longitudinal corpus of clinical narratives. *Journal of biomedical informatics*, *58*, S6-S10.

Kurosaki, S. (2013). *An analysis of the knowledge and use of English collocations by French and Japanese learners*. Universal-Publishers.

Larsen-Freeman, D. (2002). Making sense of frequency. *Studies in second language acquisition*, *24*(2), 275-285.

Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in second language acquisition*, *15*(1), 35-48.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647-672.

Lawal, A. (2003). *Stylistics in Theory and Practice.* Ilorin: Paragon Book Ltd.

Lee, C. Y., & Lin, C. C. (2013). Evaluation on second language collocational congruency with computational semantic similarity. In *PACLIC 27 Workshop on Computer-Assisted Language Learning* (pp. 534-541).

Lee, S. (2016). L1 influence on the processing of L2 collocation: An experimental study of Korean EFL learners. *Linguistic Research 33(Special Edition), 137-163.*

Leech, G. (1992). Corpora and theories of linguistic performance. *Directions in corpus linguistics*, 105-122.

Leech, G., & Rayson, P. (2014). *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.

Leśniewska, J., & Witalisz, E. (2007). Cross-linguistic influence and acceptability judgments of L2 and L1 collocations: A study of advanced Polish learners of English. *Eurosla Yearbook*, *7*(1), 27-48.

Lewis, M. (1993). *The lexical approach* (Vol. 1, p. 993). Hove: Language Teaching Publications.

Lewis, M. (1997). Pedagogical implications of the lexical approach. *Second language vocabulary acquisition: A rationale for pedagogy*, 255-270.

Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. *Perspectives on formulaic language: Acquisition and communication*, 22-46.

Lim, J. M. H. (2007). Crosslinguistic influence versus intralingual interference: A pedagogically motivated investigation into the acquisition of the present perfect. *System*, *35*(3), 368-387.

Littlemore, J. (2010). Metaphoric competence in the first and second language. *Converging Evidence in Language and Communication Research (CELCR)*, 293.

Littlemore, J., Chen, P. T., Koester, A., & Barnden, J. (2011). Difficulties in metaphor comprehension faced by international students whose first language is not English. *Applied Linguistics*, *32*(4), 408-429.

Loewen, S. & Plonsky, L. (2016). An A–Z of Applied Linguistics Research Methods. London: Palgrave McMillan.

Loewen, S., & Plonsky, L. (2015). *An A–Z of applied linguistics research methods*. Macmillan International Higher Education.

Lowenberg, P. H. (1986). Non-native varieties of English: Nativization, norms, and implications. *Studies in Second Language Acquisition*, *8*(1), 1-18.

LRC Conference (2011). "20 years of learner corpus research: looking back, moving ahead" available at: https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpus-research-2011.html [Accessed 25th of November 2018].

Lu, Y. (2016). *A Corpus Study of Collocation in Chinese Learner English*. Routledge.

Macis, M., & Schmitt, N. (2016). The figurative and polysemous nature of collocations and their place in ELT. *ELT Journal*, ccw044.

Macis, M., & Schmitt, N. (2017). Not just 'small potatoes': Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, *21*(3), 321-340.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Routledge.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press. Malden: Blackwell, 7-29.

Malinowski, B. (1923). The Problem of Meaning in Primitive Languages. In C. K. Ogden, & I. A. Richards (Eds.), The Meaning of Meaning (pp. 296-336). London: K. Paul, Trend, Trubner.

Marchand, T., & Akutsu, S. (2015). First steps in assigning proficiency to texts in a learner corpus of computer-mediated communication. *Learner corpora in language testing and assessment*, *70*, 85.

Marckwardt, A. H. (1963). English as a second language and English as a foreign language. *Publications of the Modern Language Association of America*, 25-28.

Martyńska, M. (2004). Do English language learners know collocations? *Investigationes linguisticae*, *11*, 1-12.

McCarthy, M (2014). Collocation and the Learner: wading into the depths. Available at: http://www.cambridge.org/elt/blog/2014/02/26/collocation-learner-wading-depths/ [Accessed on 10 January 2018].

McCarthy, M. (1990) *Vocabulary*. Oxford: Oxford University Press.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behaviour research methods*, *42*(2), 381-392.

McEnery, A. M., & Wilson, A. (2001). *Corpus linguistics: an introduction*. Edinburgh: Edinburgh University Press.

McEnery, T., & Wilson, A. (1996). Corpus linguistics. Edinburgh: Edinburgh University Press.

McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. *Handbook of research in second language teaching and learning*, *2*, 364-380.

McIntosh, C. (2009). Oxford collocations dictionary for student of English. Oxford University Press.

Meara, P. (1978). Learners' word associations in French. *Interlanguage Studies Bulletin*, 192-211.

Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular*, *11*(2), 29-38.

Mel'cuk, I. (1996). Lexical functions: a tool for the description of lexical relations in a lexicon. *Lexical functions in lexicography and natural language processing*, *31*, 37-102.

Mel'čuk, I. (1998). Collocations and lexical functions. In A. P Cowie (Ed) *Phraseology. Theory, analysis, and applications* (pp. 23-53). Oxford: OUP

Mel'čuk, I. A. (1974). Grammatical meanings in interlinguas for automatic translation and the concept of grammatical meaning. *Rozencvejg (hg.): Machine Translation and Applied Linguistics*, *1*.

Mel'cuk, I. A. (1974). Statistics and the relationship between the gender of French nouns and their endings. *Essays on lexical semantics*, *1*, 11-42.

Mel'čuk, I. A. (1981). Meaning-text models: A recent trend in Soviet linguistics. *Annual review of Anthropology*, *10*(1), 27-62.

Milićević, J. (2006). A short guide to the meaning-text linguistic theory. *Journal of Koralex*, *8*, 187-233.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, 211-232.

Milton, J., & Alexiou, T. (2009). Vocabulary size and the common European framework of reference for languages. In *Vocabulary studies in first and second language acquisition*(pp. 194-211). Palgrave Macmillan, London.

Montero, L., Serrano, R., & Llanes, À. (2017). The influence of learning context and age on the use of L2 communication strategies. *The Language Learning Journal*, *45*(1), 117-132.

Mora, J. K. (2001). Metalinguistic awareness as defined through research. *Available at: http://www. moramodules. com/Pages/MetalingHandout. [Accessed 26 June 2018].*

Mufwene, S. S. (2001). *The ecology of language evolution*. Cambridge University Press.

Namei, S. (2004). Bilingual lexical development: A Persian–Swedish word association study. *International Journal of Applied Linguistics*, *14*(3), 363-388.

Namvar, F. (2012). The relationship between language proficiency and use of collocation by Iranian EFL students. *3L: Language, Linguistics, Literature®*, *18*(3).

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied linguistics*, *24*(2), 223-242.

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In Sinclair, J. M. (Ed.). *How to use corpora in language teaching*. John Benjamins Publishing, *12*, 125-156.

Nesselhauf, N. (2005). Collocations in a learner corpus. *Studies in Corpus Linguistics (Vol. 14)*. Amsterdam: Benjamins.

Nizonkiza, D. (2011). The relationship between lexical competence, collocational competence, and second language proficiency. *English Text Construction*, *4*(1), 113-145.

Nizonkiza, D. (2012). Quantifying controlled productive knowledge of collocations across proficiency and word frequency levels. *Studies in Second Language Learning and Teaching*, *2*(1), 67-92.

Nizonkiza, D. (2015). Measuring receptive collocational competence across proficiency levels. *Stellenbosch Papers in Linguistics*, *44*, 125-146.

Nizonkiza, D. (2015). Measuring receptive collocational competence across proficiency levels. *Stellenbosch Papers in Linguistics*, *44*, 125-146.

Nizonkiza, D. (2015). Measuring receptive collocational competence across proficiency levels. *Stellenbosch Papers in Linguistics*, *44*, 125-146.

Nizonkiza, D., & Van de Poel, K. (2014). Teachability of collocations: The role of word frequency counts. *Southern African Linguistics and Applied Language Studies*, *32*(3), 301-316.

Nizonkiza, D., Van Dyk, T., & Louw, H. (2013). First-year university students' productive knowledge of collocations. *Stellenbosch Papers in Linguistics Plus*, *42*, 165-181.

North, B. (2014). *The CEFR in practice* (Vol. 4). Cambridge University Press.

Odumuh, A. (1987). Nigerian English. Zaria: Ahmadu Bello University Press.

Odumuh, E. (1983). *Sociolinguistics and Nigerian English*, Lagos: Sambookman.

Okoro, O. (2013). Exploring collocations in Nigerian English usage. *California Linguistic Notes*, *38*(1), 84-121.

Okunrinmeta, U. (2014). Syntactic and Lexico-Semantic Variations in Nigerian English: Implications and Challenges in the ESL Classroom. *Open Journal of Modern Linguistics*, *4*(02), 317.

Oller, J. W. (1973). Cloze tests of second language proficiency and what they measure1. *Language learning*, *23*(1), 105-118.

Osoba, G. A. (2014). English Idioms in Some Nigerian Print Media: of Norm and Deviation. *English Linguistics Research*, *3*(1), 46.

Östman, J. O. (2005). Persuasion as implicit anchoring. *Persuasion across genres: A linguistic approach*, 183.

Pajak, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1377.

Palmer, H. (1938). A grammar of English words. London: Longman.

Palmer, H. E. (1933). Second interim report on English collocations. In the Tenth Annual Conference of English Teachers under the Auspices of the Institute for Research in English Teaching. Tokyo, Institute for Research in English Teaching, 1933.

Pavlenko, A. (2003). Eyewitness memory in late bilinguals: Evidence for discursive relativity. *International Journal of Bilingualism*, *7*(3), 257-281.

Pavlenko, A. (2008a) Structural and conceptual equivalence in the acquisition and use of emotion words in a second language. *The Mental Lexicon* 3 (1), 91-120.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. *The bilingual mental lexicon: Interdisciplinary approaches*, 125-160.
Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. *The bilingual mental lexicon: Interdisciplinary approaches*, 125-160.
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection

Pecina, P. (2010). Lexical association measures and collocation extraction. *Language resources and evaluation*, *44*(1-2), 137-158.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, *18*(1), 75-94.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113-138.

Philip, G. (2011). *Colouring meaning: Collocation and connotation in figurative language*. John Benjamins Publishing.

Phoocharoensil, S. (2011). Collocational errors in EFL learners' interlanguage. *Journal of Education and Practice*, *2*(3), 103-120.

Phoocharoensil, S. (2012). Cross-linguistic influence: Its impact on L2 English collocation production. *English Language Teaching*, *6*(1), 1.

Phoocharoensil, S. (2014). Exploring Learners. Developing L2 Collocational Competence. *Theory and practice in language studies*, *4*(12), 2533.

Platt, J. T., & Weber, H. (1980). *English in Singapore and Malaysia: Status, features, functions*. Oxford University Press.

Polit, D. F., Beck, C. T., & Hungler, B. P. (2001). *Essential of Nursing Research, methods, appraisal and utilization*. (5th ed.). Philadelphia: Lippingcott Williams & Wilkins.

Porzig, W. (1934). Wesenhafte Bedeutungsbeziehungen. Beträge zur Geschichte der deutsche Sprache und Literatur, 58, 70-97.

Potter, M. C., So, K. F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of verbal learning and verbal behavior*, *23*(1), 23-38.

Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, *26*(1), 8-14.

Read, J. (2000). *Assessing vocabulary* (Cambridge language assessment series). Cambridge: Cambridge University Press.

Revier, R. L. (2009). Evaluating a new test of whole English collocations. In A. Barfield & H. Gyllstad (Eds.), Researching collocations in another language: Multiple interpretations (pp. 125-138). Basingstoke: Palgrave Macmillan.

Ridha, N. S. A., & Al-Riyahi, A. A. (2011). Lexical collocational errors in the writings of Iraqi EFL learners. *ADAB AL-BASRAH*, (58), 24-51.

Roy, B. C., Frank, M. C & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. In Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society, July 29 – August 1, 2009, Vrije Universiteit, Amsterdam, Netherlands.

Salim, B. (2001). *A companion to teaching of English*. New Delhi: Atlantic Publishers & Dist. Ltd.

Saywitz, K., & Cherry-Wilkinson, L. (1982). Age-related differences in metalinguistic awareness. In S. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought and culture.* Hillsdale, NJ: Erlbaum.

Schiebert, W. (2009). Corpus Linguistics: Lexicography and Semantics: Introduction to Concordance and Collocations. Munich: GRIN Verlag.

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied linguistics*, *11*(2), 129-158.

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge University Press.

Scott, M. (1999). WordSmith Tools users help file. Oxford: Oxford University Press.

Seidlhofer, B. (2006). English as a lingua franca – so what's new? Available at: www.anglistik.uni-halle.de/anglistentag2006/abstractVaktuell.pdf [Accessed 3rd of March 2018]

Seretan, V. (2008). *Collocation extraction based on syntactic parsing* (Doctoral dissertation, Ph. D. thesis, University of Geneva).

Seretan, V. (2011). *Syntax-based collocation extraction* (Vol. 44). Heidelberg: Springer Science & Business Media.

Shehata, A. K. (2008). *L1 Influence on the reception and production of collocations by advanced ESL/EFL Arabic learners of English* (Doctoral dissertation, Ohio University).

Shei, C. C. (1999). A brief review of English verb-noun collocation. *Available on-line at http://www. dai. ed. ac. uk/homes/shei/survey. html*. [Accessed 24 June 2017].

Shih, R. H. H. (2000). Collocation deficiency in a learner corpus of English: From an overuse perspective. In *Proceedings of the 14th Pacific Asia Conference on Language, information and Computation* (pp. 281-288).

Shitu, F. (2015). Collocation Errors in English as Second Language (ESL) Essay Writing. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, *9*(9), 3176-3183.

Shitu, F. M. (2015). Collocation Errors in English as Second Language (ESL) Essay Writing. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, *9*(9), 3176-3183.

Siepmann, D. (2005). Collocation, colligation and encoding dictionaries. Part I: Lexicological aspects. *International Journal of Lexicography*, *18*(4), 409-443.

Simon, M. K. (2011). *Dissertation and Scholarly Research: Recipes for success.* Seattle, WA: Dissertation Success, LLC.

Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford: Oxford University Press.

Sinclair, J. (2004b). Trust the text: Language, corpus and discourse. London: Routledge.

Jones, S. & Sinclair, J. (1974). English lexical collocations. *Cahiers de Lexicologie*, 24, 15-61.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*(3), 429-458.

Smadja, F. (1993). Retrieving collocations from text: Xtract. *Computational linguistics*, *19*(1), 143-177.

Smiskova, H., & Verspoor, M. (2017) Development of chunks in Dutch L2 learners of English. In Tribushinina et al (2017) *Usage-based approaches to language acquisition and language teaching. Berlin: Mouton de Gruyter*.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121-159.

Song, L., & Wolter, B. (2017). Effects of L1 Transfer on L2 Learners' VN Collocational Use: A Corpus-based Study from Semantic Preference and Semantic Prosody Perspective.

Available at: https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper55.pdf [Accessed 24th of November 2018].

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of language*, *2*(1), 23-55.

Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Oxford: Blackwell.

Taiwo, R. (2001). Lexico-semantic relations errors in senior secondary school students' writing. *Nordic Journal of African Studies*, *10*(3), 366-373.

Taiwo, R. (2004). Helping ESL learners to minimize collocational errors. *The Internet TESL Journal*, *10*(4), 2004.

Taiwo, R. (2010). *Collocation in non-native English: A study of Nigerian ESL writing*. Obafemi Awolowo University, Nigeria.

Talakoob, F., & Koosha, M. (2017). Productive and Receptive Collocational Knowledge of Iranian EFL Learners at Different Proficiency Levels. *International Journal of Applied Linguistics and English Literature*, *6*(7), 11-16.

Tarp, S. (2011). Pedagogical lexicography: Towards a new and strict typology corresponding to the present state-of-the-art. *Lexikos*, *21*.

Thomas, J. (1988). The role played by metalinguistic awareness in second and third language learning. *Journal of Multilingual & Multicultural Development*, *9*(3), 235-246.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Harvard University Press

Tomasello, M. (2009). The usage-based theory of language acquisition. In *The Cambridge handbook of child language* (pp. 69-87). Cambridge Univ. Press.

Tono, Y. (2003). Learner corpora: design, development and applications. In *Proceedings of the Corpus Linguistics 2003 conference* (pp. 800-809). Lancaster: University Centre for Computer Corpus Research on Language.

Trager, G. L. (1940). The Russian gender categories. Language, 300-307. Baltimore, MD: Linguistic Society of America.

Trantescu, A. M. (2015). Conceptual Motivation of English and Romanian Shoulder, Arm and Hand Idioms. A Contrastive Approach. *EUROPEAN LANDMARKS OF IDENTITY*, 53.

Tribushinina, E., Valcheva, E., & Gagarina, N. (2017). 9 Acquisition of additive connectives by Russian-German bilinguals: A usage-based approach. *Usage-Based Approaches to Language Acquisition and Language Teaching*, *55*, 207.

Van der Meer, G. (1998, August). Collocations as one particular type of conventional word combinations: Their definition and character. In *Proceedings of the 8th Euralex Conference* (pp. 4-8).

Van der Wouden, T. (1997). *Negative contexts: Collocation, polarity and multiple negation*. New York: Routledge.

Van Kleeck, A. (1982). The emergence of linguistic awareness: A cognitive framework. *Merrill-Palmer Quarterly (1982)*, 237-265.

Van Rooy, B., & Schäfer, L. (2002). The effect of learner errors on POS tag errors during automatic POS tagging. *Southern African Linguistics and Applied Language Studies*, *20*(4), 325-335.

Vivian de Klerk. (2006). The features of 'teacher talk' in a corpus-based study of Xhosa English. *Language Matters* 37:2, 125-140.

Vulchanova, M., Vulchanov, V., & Stankova, M. (2011). Idiom comprehension in the first language: a developmental study. *Vigo International Journal of Applied Linguistics*, *8*.

Vyatkina, N. (2015). Corpus-Driven Instruction. University of Oregon: Centre for Applied Second Language Studies (CASLS). http://caslsintercom.uoregon.edu/content/18926. [Accessed 28 May 2018].

Waggoner, J. E., Palermo, D. S., & Kirsh, S. J. (1997). Bouncing bubbles can pop: Contextual sensitivity in children's metaphor comprehension. *Metaphor and Symbol*, *12*(4), 217-229.

Waller, T. (1993). Characteristics of near-native proficiency in writing. *Near-native proficiency in English*, *2*, 183-293.

Walqui, A. (2000). Contextual Factors in Second Language Acquisition. ERIC Digest. Available at: https://files.eric.ed.gov/fulltext/ED444381.pdf [Accessed 30 June 2017]

Wang, Y., & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. ICAME Journal, 32, 201-232.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*, *28*(1), 46-65.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*(1), 91-120.

Winner, E., Rosenstiel, A., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, 12, 289-297.

Wiśniewska-Kin, M. (2017). Children's Metaphor Comprehension and Production. *Stanisław Juszczyk*, 87.

Wisniewski, K. (2017). Empirical learner language and the levels of the common European framework of reference. *Language Learning*, *67*(S1), 232-253.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*(4), 430-449.

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *36*(5), 1193-1221.

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231-254.

Xiao, Z., & McEnery, A. (2005). Two approaches to genre analysis: Three genres in modern American English. *Journal of English Linguistics*, *33*(1), 62-82.

Xu, Q. (2016). Item-based foreign language learning of give ditransitive constructions: Evidence from corpus research. *System*, *63*, 65-76.

Yamashita, J. & Jiang N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*(4), 647–668.

Yumanee, C., & Phoocharoensil, S. (2013). Analysis of collocational errors of Thai EFL students. *LEARN Journal: Language Education and Acquisition Research Network*, *6*(1), 88-98.

Zareva, A. (2007). Structure of the second language mental lexicon: How does it compare to native speakers' lexical organization? *Second language research*, *23*(2), 123-153.

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, *27*(4), 567-595.

ZHANG, X., & Nannan, L. I. U. (2014). Exploring the Second Language Mental Lexicon with Word Association Tests. *Cross-Cultural Communication*, *10*(4), 143-148.

Zibin, A. (2016). The comprehension of metaphorical expressions by Jordanian EFL learners. *SAGE Open*, *6*(2), 2158244016643144.

Zibin, A., & Hamdan, J. (2014). The acquisition of metaphorical expressions by Jordanian EFL learners: A cognitive approach. *Saarbrücken: Lambert Academic Publishing*.

Zobl, H. (1980a). Developmental and transfer errors: their common bases and (possibly) differential effects on subsequent learning. *Tesol Quarterly*, 469-479.

Zobl, H. (1980b). The formal and developmental selectivity of LI influence on L2 acquisition. *Language learning*, *30*(1), 43-57.

Žolkovskij, A. K., & Mel'čuk, I. A. (1967). On semantic synthesis (of texts). *Russian. Probl. Cybern*, *19*, 177-238.

Appendix A- Pilot Study Cloze Test

Carefully read the passage below and fill in the gaps with the word or phrase you consider most appropriate in this context. ***DO NOT*** *consult a dictionary or any reference material. This is* ***NOT*** *a test of your intelligence; the test is purely for academic research purpose.* ***ALL RESPONDENTS*** *would be treated as* ***ANONYMOUS.*** ***Complete this test within 1hr***

Background Information

1. What is your highest qualification? ……...
2. Do you have a credit pass in English Language in your Nigerian

   WEASC/SSCE/GCE? ……
3. What is your discipline? …….
4. What is your first language? ……...
5. Do you use English language regularly at work? ……....

THE MEMOIR OF A YOUNGSTER

I never knew hunger when I was growing up in rural Nigeria. We had all the food we wanted. All the food was fresh, and we wasted awful lot of food because we had no refrigerator to preserve them. Dad was very hard working; he was the father who was capable ……..… providing for his home. Though he was friendly, he wouldn't take kindly to any child failing to comply ……..….. his rules. It was like there was this unwritten constitution which we must all adhere …… Just like Dad, Mum would always insist …….…… doing the right thing and always well-behaved. She was very

interested ……..…. our welfare. When we had all gone to bed, she would come around to see if we were all well covered by the mosquito net. My mother was always conscious ……..…. the deadly effect of mosquito bites. But it didn't matter how well we were covered, the invading mosquito would always find some ways of feeding on our precious blood and infecting us with malaria virus. That meant we had to..... frequent visits to the local dispensary. Sadly, some kids in our village didn't survive the malaria attacks.

Our house was like every typical village house in Southwest Nigeria. It was made of bricks and thatched roof, with tiny window – barely wide enough to let fresh air into our rooms; and large part of the wall was darkened by smoke from the kitchen. Here in the village, social life was non-existent – no TV set, not telephone, even a wall clock was a luxury. The only thing we had in abundance was food – this wasn't the case for most families who lived in …….. poverty. Our parents' focus wasn't to amass wealth, but just to have enough to …….. the need of the family. Most of the time, they concentrated ……..…… providing food for the kids and we always had enough of that. But it wasn't all about food. I have never seen a mother who was so mindful of safety and security as my mother! She wouldn't allow us to play with any object with which we could …… a wound on ourselves.

My father would ……. music which he would sing for us whenever we gathered in the bright moonlight before going to bed. His music, most of the time, was to teach morals. He would sing about men who …… bravery in time of war, young people who ……. temptation to steal when they had the opportunity to do so and thereby earning a good name. He would …… a story of a …… of lions that used to roam the forest before uncontrolled timber cutting destroyed their habitat. He told of a time when the villagers mistook a ……. of whales that often came close to the shore for the colonial masters' submarines. We didn't enjoy that story because we were too young to know what whales and submarines were. It only left us wondering what a mysterious creature my Dad was talking about. In fact, I thought submarines were some rare species of marine mammals.

Before ending the moonlight storytelling and singing, he would ask us a few questions to …… our attention to the morals of the stories, and then we would

…… a prayer before going to bed. So, we grew up to know our father not just as the food provider, but as an entertainer as well. If someone had asked me to

…… a candidate for Best Dad's Award, my father would be my candidate!

In addition to having abundant food, we also had peace in abundance at home. Mum and Dad were a perfect match! You wouldn't see them arguing …… on any issue. This was in ……. contrast to my neighbour's parents. Their father wouldn't consult their mother before ……. crucial decisions. This would often result ……. frequent quarrels. Unfortunately, the pair couldn't get along well; they had irreconcilable difference and a customary court had to …… their marriage. The court ordered the husband to vacate the family home. But the arrogant father wouldn't accept such, in his view, humiliating verdict; he would rather destroy the family house than see his estranged wife live in it with the kids. About three months after their divorce, the father who had been spying ……. the mother, came back to set the house on fire. As the fire was burning, the mother ran into the burning house with …….. abandon to rescue her youngest child who was sleeping in the house at the time. But it was too late; the fire had spread quickly killing the child and leaving the mother severely burned. There was much grief in the neighbourhood. The father was arrested and charged …… arson and manslaughter. During his trial, the jury didn't take long to ……. a verdict.

He was convicted …… arson and manslaughter and sentenced …… 27 years imprisonment. As you would expect, he appealed …… the ruling but the high court …… his appeal, insisting 27 years was appropriate to deter others from

……. the law. What a tragic end!

We resumed our moonlight story telling after about two months break following this incident. As I was about to start my primary education, my Dad's stories focused on the importance of good education. I was uncomfortable with the idea of leaving home for a boarding school, but my father's stories …… my fear and prepared my mind to

adapt …….. this change. I left for school a day before school officially resumed because I had to travel a long distance. It was an entirely new experience living in the dormitory and queuing for food. I hated staying on the queue for long; I sometimes wished I could just ……. the queue and get my food before other pupils. It didn't take long to get used to my new environment; thanks to my father's stories. Right from my first night in the hostel, I started …… a diary**.** I wanted to make sure I had some stories for my Dad at the end of the school year. Though I wasn't lonely because I had…….. new friends, I still missed my mother; I missed her food more! I missed my Dad and my siblings, too. I would give them a

…….. of flowers when I return home for holiday.

Appendix B – CERF Self-Assessment Grid

	A1	A2	B1	B2	C1	C2
Listening	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.
Reading	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.
Spoken Interaction	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).	I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.	I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.
Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.	I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.	I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.
Writing	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages relating to matters in areas of immediate needs. I can write a very simple personal letter, for example thanking someone for something.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.	I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.	I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.