

**A LEXICAL EXAMINATION OF THE COLLOCATION PATTERNS AND  
DEGREES OF OPACITY IN MULTIWORD UNITS: ILLUSTRATIONS  
FROM ANIMAL FARM**

**By**

**MBUTU ANTHONY MAINA**

**A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT FOR THE  
AWARD OF THE DEGREE OF MASTER OF ARTS, IN ENGLISH AND  
LINGUISTICS IN THE SCHOOL OF HUMANITIES AND SOCIAL  
SCIENCES, KENYATTA UNIVERSITY.**

**KENYATTA UNIVERSITY LIBRARY**

**DECEMBER 2008**

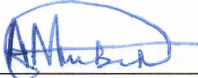
Mbutu, Anthony Maina  
*A lexical examination  
of the collocation*



2013/421629

**DECLARATION**

This dissertation is my original work and has not been presented for a degree in any other university.

Signed: 

Date: 6/12/08

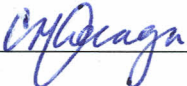
**Mbutu A. Maina**

We as the university supervisors, confirm that the work reported in this dissertation was carried out by the candidate. This dissertation has been submitted for examination with our approval as university supervisors.

Signed: 

Date: 16.12.2008

**Dr. Ruth W. Ndung'u**  
Department of English and Linguistics  
Kenyatta University

Signed: 

Date: 8/12/2008

**Mr. C.M. Gecaga**  
Department of English and Linguistics  
Kenyatta University

## DEDICATION

This dissertation is dedicated to my dear parents Hezron Mbutu Wahome and Jedidah Wairimu Mbutu, who sacrificed their all to ensure that I obtained an education. Mom taught me to pray to God before attempting any examination while Dad always purchased the text books I needed which always bore the affectionate designation 'My son.' God bless you.

I also dedicate it to my wife Lucy Wambui, My son Zadok, my daughter Wairimu in addition to Britney, Hezron, Mercy, Lucy, John, Irene and Jack.

## ACKNOWLEDGEMENTS

I would like to take this opportunity to pay tribute to the following for their invaluable support in the course of the research.

First, I thank my Lord and saviour Jesus Christ for His enabling grace without which I would have given up. Indeed the completion of this work confirms the truth of the words: “Unless the Lord builds the house, they labour in vain who build it...” (Psalms 127:1. It is His provision, protection and inspiration that has brought me this far.

Secondly, I would like to appreciate the effort of my dear wife Lucy Wambui who daily ensured I was ready to vanquish each day’s challenges. May God remember her in a mighty way for her selflessness. My son Zadok Mbutu and daughter Wairimu Maina also deserve special mention for giving me the impetus to succeed, as I thought about them day by day. I also wish to acknowledge my parents Hezron Mbutu and Jedidah Wairimu for their advice and guidance. I also extend my gratitude to my siblings Jackson Ngari, John Mithamo, Lucy Wambui and Irene Wanjiru for encouraging me when I was on the verge of despair-may the almighty God repay you in abundance. I appreciate my nephew Hezron Mbutu and niece Britney Wairimu whose cheerful smiles always invigorated me.

In addition, I appreciate the supervision of Dr. Ruth Wangechi Ndung'u and Mr. C. M. Gecaga for their guidance in the course of the study. My sincere prayer is that God may grant them the desires of their hearts. I am grateful to Dr. Ndung'u for the countless times she lent me her textbooks and various other resources on lexis. May she forever reap the fruits of her generosity. The other lecturers in the department of English also provided guidance at one time or the other and for that, too I am very grateful.

The members of the non-teaching staff in the Department of English and my fellow students all provided spiritual, moral and material assistance may God remember every one of them.

My appreciation also goes to Mr. Waboiya, the principal Gatwe Secondary School in Kirinyaga District also deserves special mention for recommending my study leave application to the Teachers' Service Commission even though I was the only English language teacher in the school-may God remember him for not standing in my way when he could have done so. I also extend my thanks to pastor Moses Kamau of Kibirigwi Redeemed Gospel Church for standing by me in prayer then and now. May God enlarge his territory in all aspects of his life.

Finally, I appreciate brother Mugambi of the Kenyatta University Christian Post Graduate Fellowship for the way he assisted me with the formatting of the dissertation. May God bless him abundantly for being such a blessing to me.

## ABSTRACT

The aim of this study is to lexically examine the types of opacity and collocation patterning in multiword units with illustrations from, *Animal Farm* (2002), edition. The objectives of the study are: to identify and describe the types of multiword units used in *Animal Farm*, to identify and describe the patterns of collocation in the multiword units; and to find out and describe the degrees of opacity in the multiword units.

The research design used is the case study design. In addition, the research identifies multiword units from the whole novel, which ensures that, the frequencies generated from the multiword units are representative of the whole text. Intensive reading of the novel is done to identify and extract the multiword expressions. The next thing the analyses do is to sample ten units in each of the four categories of the data for a detailed analysis. Forty units are sampled in total but since not all the multiword units have the aspects the study seeks, the analyses select 8 multiword units with an occurrence of six times and above from the sample and subjects them to the restriction cline. The analyses also select 18 multiword units, which are randomly chosen and subject them to the opacity cline. The data generated are multiword units, which are classified into phrasal verbs, compound words, prefabs and idioms. There are data on types of multiword units, data on patterns of collocation and data on degrees of opacity.

The Lexical Approach Theory posited by (Lewis, 1993) and the Idiom Principle, which was propounded by (Sinclair, 1991), guide the study. The two theories help account for the multiword units in the novel and aid in the analysis of the data. Quantitative and qualitative techniques are used to analyze the data. The categories or types of multiword units are presented and analyzed using percentages, graphs and charts. The degree of opacity and patterns of collocation are analyzed using content analysis. All the multiword units extracted from the novel are subjected to graphical analysis. The researcher then purposively samples ten multiword units from each category and further analyzes them using the opacity and restriction clines. The data analyses are then presented and recommendations for further research made. The following are some of the aspects, which the findings reveal about multiword units.

First, Multiword units can be ranged along restriction and opacity clines. Secondly, multiword units are examples of complex collocations. In addition, there are four types of multiword units. Another aspect is that culture affects the interpretability of multiword units. Finally, patterns of collocation in the multiword units affect the meanings of the multiword units.

These findings are useful to educationists and linguists because they dictate the types of ESL learning resources which should be developed, the teaching and learning strategies that should be used as well as design of the language syllabus which is lexical in nature. The explanation behind this proposition is that the study underscores the central place of lexis in language as opposed to the traditional scenario where lexis was relegated to the backyard.

## TABLE OF CONTENTS

Title Page.....	i
Declaration.....	ii
Dedication.....	iii
Acknowledgements.....	iv
Abstract.....	vii
Table of Content.....	ix
List of Tables and Figures .....	ixii
Definition of Terms .....	xiv
<b>CHAPTER ONE: INTRODUCTION.....</b>	<b>1</b>
1.1 Background to the study.....	1
1.2 Statement of the problem.....	5
1.3 Objectives of the study .....	6
1.4 Research questions.....	7
1.5 Research assumptions.....	7
1.6 Rationale of the study .....	7
1.7 Scope and Limitation.....	10
<b>CHAPTER TWO: LITERATURE REVIEW AND THEORETICAL CONSTRUCT.....</b>	<b>12</b>
2.1 Introduction.....	12
2.1.1 Lexis.....	12

2.1.2	Multiword units.....	14
2.1.3	Language and literature.....	22
2.2	Theoretical framework.....	24
CHAPTER THREE: RESEARCH METHODOLOGY.....		28
3.1	Research design.....	28
3.2	Sampling techniques and sample size.....	28
3.3	Data collection.....	30
3.4	Data analysis.....	30
CHAPTER FOUR: DATA ANALYSIS AND PRESENTATION.....		33
4.0	Introduction.....	33
4.1	The multiword units in the data.....	34
4.1.1	Compound words.....	35
4.1.2	Phrasal verbs.....	36
4.1.3	Idioms.....	38
4.1.4	Prefabs.....	39
4.1.5	Conclusion.....	39
4.2	The restriction and opacity clines.....	41
4.2.1	Collocation restriction in the multiword units.....	44
4.2.2	Unrestricted collocations.....	45
4.2.3	Semi-restricted collocations.....	49
4.2.4	Familiar collocations.....	50
4.2.5	Restricted collocations .....	54

4.3 Opacity in the multiword units .....	56
4.3.1 Transparent multiword units.....	57
4.3.2 Semi-idioms.....	58
4.3.4 Semi-transparent collocations .....	61
4.3.4 Opaque collocations.....	63
4.4 Relationship between restriction and opacity.....	68
CHAPTER FIVE: SUMMARY OF FINDINGS AND RECOMMENDATIONS.	72
5.0 Introduction.....	72
5.1 Summary of findings .....	72
5.1.1 Types of multiword units.....	72
5.1.2 The collocation restriction in the multiword units.....	73
5.1.3 Opacity in the multiword units .....	74
5.2 Recommendations.....	78
5.3 Areas for further research .....	79
5.4 Conclusion .....	80
BIBLIOGRAPHY .....	82
APPENDICES.....	86
Appendix 1: Compound words identified in the data .....	86
Appendix 2: Phrasal verbs identified in the data.....	88
Appendix 3: The idioms identified in the data.....	89
Appendix 4: The prefabs identified in the data.....	92

Appendix 5: The multiword units sampled from each of the four classes of multiword units.....95

Appendix 6: The multiword units that were subjected to the restriction/opacity clines .....96

**LIST OF TABLES AND FIGURES**

Table 3.1 The semantic opacity cline for the multiword units.....	31
Table 3.2 The restriction cline for the patterns of multiword units.....	32
Table 4.1 The distribution of different multiword units in the data.....	36
Table 4.2: The distribution of compound words in the data.....	40
Fig 4.1 an exploded pie chart showing the distribution of the various types of multiword units in the data.....	41

## DEFINITION OF TERMS

**Corpus** -A body of naturally occurring language

**Lexis** - The vocabulary of a language in contrast to its grammar

**Lexicon** -The set of all the words and idioms of any language

**Multiword units** -A unit of meaning consisting of two or more words whose grammatical form and lexical content is wholly or largely fixed.

**Collocation** - The way in which words are used together regularly

**Transparency**-Aspect of multiword units in which the meaning can be inferred from the constituents.

**Opacity** -The case where prior encounter is necessary for one to know the meaning of a multiword unit

**Cline** -Scale that shows how multiword units may be classified from the most opaque to the most transparent.

**Lexicology** -The study of vocabulary items

**Lexeme** -The smallest unit in the system of meaning of lexical works. It can be distinguished from other similar units in lexical research.

**Lexicography** -Compiling of dictionary

**Lexical** -Pertaining to words.

## CHAPTER ONE

### INTRODUCTION

#### 1.1 Background to the Study

Kristen Malmkjer (1991) defines lexis as the study of vocabulary of all languages in all its aspects: For example, words and their meanings. There are two categories of words; grammatical and lexical words. Grammatical words are also referred to as functional words while lexical words are known as content words (Carter, 1998). There are several concerns of lexis. To begin with, lexis explores the meaning of words, which is also known as semantics-both content, and grammatical words may convey meaning according to lexicologists. Secondly, lexis deals with how words relate to one another. It is the view of modern linguists that words do not occur in isolation but that they form associations with one another in any given context (Singleton, 2000). These associations are of interest to linguists and are referred to as collocations. Another point of concern to lexicologists is how words combine with one another and the relationships between vocabulary and other areas of description of language such as phonology, morphology and syntax. In other words, lexis refers to the study of lexical items, which may be multiword unit such as proverbs or single words such as tree. The three concerns of lexis discussed above are the same thing in that words relate to one another in various ways to express meaning, which is in turn affected by the

sound system, the word formation processes as well as the syntactic considerations with regard to a given language.

A multiword item is a vocabulary item, which consists of a sequence of two or more words (a word being simply an orthographic unit) for instance *knock off*. The sequence of words semantically and or syntactically forms a meaningful and inseparable unit. If one for example considers an item such as good morning, there is no way its constituents can be separated and still express the intended meaning (Schmitt and McCarthy, 1997). Multiword units are examples of collocation in a spoken or written text, where collocation means the way in which some words are often used together (Longman advanced English dictionary). Singleton (2000) defines collocation as the way in which certain words habitually keep company.

All languages comprise of single and multiword units and the latter have been the focus of language research as evidenced by the works of Singleton (2000), Sinclair (2004) among others. In fact, some lexicologists such as Lewis (1997) assert that language does not consist of the traditional grammar versus vocabulary distinction but often of multiword prefabricated chunks. This assertion explains why the study of multiword units is important in our time. Other scholars such as Lewis (1993) have proposed the scrapping of the traditional grammar versus-lexis dichotomy and this has given lexis a central place in linguistics.

As Schmitt (1997) observes, multiword units are used extensively; by both native and non-native users of any language. The usage traverses both written and spoken language. Lewis (1993) further argues that, much of language comprises of multiword prefabricated chunks, which necessitates the need for exploratory research with regard to multiword units. A good reservoir of multiword units also improves the quality of one's writing according to Howarth (2007). This implies that every speaker of a language must endeavor to improve their command of multiword units since they are so crucial. Conklin and Schmitt (2007) also observe that at least one third to one-half of language is composed of formulaic elements. The formulaic elements may also be used in a variety of ways. There has been much research in formulaic language but this has largely been in written discourse.

From the foregoing, it may be concluded that multiword units are encountered all the time in the course of people's interaction with language; be it written or spoken. It is against this background that this research focuses on George Orwell's novel *Animal Farm*, a text that has numerous multiword units. In fact, anything from a third to a half of the novel is primarily: multiword units, if Schmitt and Conklin's (2007) argument is anything to go by. Other studies by Schmitt and Conklin (2007) have shown that the processing advantage of formulaic expressions is faster for both native and non-native users of a language compared to non-formulaic ones. Hence, there is a possibility that non-native users of a language can enjoy the same processing advantage as the native speakers provided

they are able to concentrate on the multiword units and manipulate them as expected.

*Animal Farm* has been studied widely from the perspective of literature as evidenced by the numerous commentaries based on the text such as: Calder (1987), which is a study guide to George Orwell's *Animal Farm* and *Nineteen Eighty-Four*. The work discusses the stylistic elements as well as the thematic concerns of the novels. The present study can also be categorized as a stylistic analysis since it is an in depth investigation of the language of a literary text: *Animal Farm*.

The present study examines the types of multiword units in the novel, their degree of opacity as well as the patterns of collocation in the multiword units on the basis of the above premise. Multiword units are an indicator of proficiency (Gairns and Redman, 1990) that is why it was deemed necessary to identify and discuss in depth the types of multiword units in the text. The various categories of multiword units are: compound words, phrasal verbs, idioms and prefabs. There is need therefore; to not only explore the diversity of multiword units but to also make recommendations to language researchers and practitioners on how to enhance their acquisition.

## 1.2 Statement of the Problem

This study considers in totality the collocation patterns and degrees of opacity in multiword units, which are, found in all languages. Multiword units are also called formulaic expressions and they are in many respects extreme cases of fixed collocations according to Schmitt and McCarthy (1997).

There are many different types of multiword items such as: idioms, prefabs, phrasal verbs and compound words. It is true that, multiword units may either result from grammatical constructions or lexical / semantic processes of fossilization however; the present study is only interested in the latter since as Schmitt (1997) says these are the bona fide multiword units otherwise referred to as the stable collocations. Opacity is the degree to which the meaning of a multiword unit is openly recoverable from its constituents. It is a problem to non-native speakers of a language as they are unable to interpret them correctly. This calls for exposure to multiword units sufficiently so that the problem may be overcome.

Multiword units, pose problems to non-native users of any language. These problems are like errors resulting from improper collocations (Redman and Gairns, 1990) as well as using formulaic expressions in wrong contexts (Schmitt, 1997). Even the native users of a language vary in their command of multiword units,

which implies that knowledge, and proper manipulation of multiword units is essential to all people no matter what their station in life is.

There has been great interest in multiword and other collocation phenomena (Sinclair, 2004). The novel simply aids the researcher to validate the claims that have been made concerning multiword units a novel is an ideal source of multiword units since the language therein is real time or natural such that one examines the units as they are actually used by the author. A novel written by a native speaker is appropriate for the study since Collie and Slater (1992) observe that, although the literary world of novels and plays is a created one, it offers a full and vivid context in which characters from many social backgrounds are depicted. This offers an indirect route to understanding how language is used in the culture of origin. Collie and Slater also assert that languages are culture specific hence; the need to focus on a text by a native speaker of English, which best illustrates how the native speakers manipulate compound words, phrasal verbs idioms and prefabs.

### **1.3 Objectives of the Study**

This study has the following objectives:

1. To identify and describe the types of multiword units used in *Animal Farm*.
2. To find out and to describe the degree of opacity in the multiword units.

3. To explore and describe the patterns of collocation in the multiword units

#### **1.4 Research Questions**

This study has the following questions

1. What types of multiword units have been used in *Animal Farm*?
2. What is the degree of opacity in the multiword units?
3. What patterns of collocation are inherent in the multiword units?

#### **1.5 Research Assumptions**

This study is based on the following assumptions:

1. Different types of multiword units are used in *Animal Farm*.
2. The multiword units in *Animal Farm* form an opacity cline.
3. The multiword units in *Animal Farm* have varying collocation patterns.

#### **1.6 Rationale of the Study**

Language research is undergoing radical change geared towards greater efficiency- of particular importance is the central role accorded Lexis today in linguistic circles as Lewis (1993) observes. The Lexical Approach postulates that language comprises of multiword prefabricated chunks hence there is need for second and foreign language users of any language to concentrate on these chunks. The ability to use the multiword units or chunks is the mark of proficiency according to Schmitt (1997). Greater emphasis has been given to the understanding of the nature of lexis in naturally occurring language. With this in mind, an examination

of possible sources of these chunks for foreign and second language users is of utmost importance.

Multiword units have implications in second language research. This is because the lexical approach to language requires the presentation of real time data. Novels are excellent sources of multiword unit, since according to Conklin and Schmitt (2007) about one-third to one-half of language is formulaic sequences. Exposure to novels would, therefore; enable users of any language to interact with multiword units in ideal contexts of usage. This in turn would lead to faster comprehension of language, as the chunks are the pillars of any language. It should also be noted that: multiword units are a phenomenon of all languages hence, anyone who reads material written by a native speaker must understand the concept of multiword units.

George Orwell's novel, *Animal Farm*, has been examined for various aspects especially in the field of literature; there are numerous commentaries based on *Animal Farm* such as Calder (1987), Wykes (1987) and Oxley (1976). However, none of these researches seeks to find out the instances of multiword units in the text. Therefore, the findings of this study for sure expand our linguistic repertoire as they prove what many a linguist has always stated: language mainly comprises of prefabricated chunks or multiword units for that matter. The choice of a literary text is occasioned by the fact that literature is language used creatively, literary

texts therefore, mirror syntactic and grammatical aspects of a language except where authors use deviations and foregrounding to achieve certain effects (Crystal, 1969).

Moreover, linguistic components are core to the understanding of the literary message, according to Collie and Slater (1995). *Animal Farm* is also an easy to read text that depicts a scenario many of us can identify with namely, betrayal of trust and abuse of power. Moreover; a native speaker of English has written it; which implies it is a true reflection of a “pure” English lexicon. As such, the novel could be a source of real time data to which language users may be exposed for faster acquisition of multiword units.

According to Schmitt (2004), there is growing evidence that native speakers of, language tend to store and utilize multiword items more productively than non - native speakers utilize. Liao and Fukuya (2003) assert that even fairly advanced second language speakers avoid using multiword items. This explains why a text written by a native speaker best illustrates the concept of multiword items. Since language is culture specific (Collie and Slater, 1995) a native speaker’s lexicon provides an ideal corpus to build up the vocabulary stock of non-native users. Collie and Slater also argue that novels provide an indirect route to understanding how language is used in the country of origin. It gives one insight since language and culture are inextricably linked.

## 1.7 Scope and Limitation

This study is limited to a lexical description of the multiword items in George Orwell's novel, *Animal Farm*. Other aspects that could be studied are stylistics, cohesion, sentence patterns and lexical density. These aspects have not been studied, as they fall outside the scope of the present research. Consequently, the researcher only focuses on multiword units; which are examples of extreme collocations as incorporating any other aspect would make the research too broad. This research only considers the stable collocations resulting from lexical and semantic fossilization, and leaves out the unstable ones such as those resulting from grammatical rules, since according to Schmitt (1997) the latter are NOT collocations. Moreover, the types of multiword units the researcher uses are compound words, phrasal verbs, idioms and prefabs (Schmitt, 1997). Needless, to say, there are other categorizations such as those of Singleton (2000), Carter (1998) among others. Nevertheless, Schmitt's approach is sufficient as far as this research is concerned.

The research is also restricted to George Orwell's novel *Animal Farm* since any one novel is rich in linguistic data; an additional text was therefore unnecessary. In addition, the research is constrained by the fact that, a lexical research requires intensive reading; it is therefore impossible to establish the occurrence of multiword units effectively, with more than one literary text.

The present research is library based, which means the researcher does not investigate how the multiword units are taught. Recommendations on how teachers may source teaching material to boost mastery of formulaic sequences are however; made based on the findings

## CHAPTER TWO

### LITERATURE REVIEW AND THEORETICAL CONSTRUCT

#### 2.1 Introduction

This section puts the present study in the perspective of related works. First, there is a review of general works on lexis, then a review of literature on multiword units, followed by a brief mention of the connection between language and literature. Finally, there is a discussion of the theoretical framework the study will adopt.

##### 2.1.1 Lexis

Lexis is defined as the vocabulary of a language in contrast to its grammar or syntax according to Richards (1987). The Oxford English dictionary defines lexis as the diction or wording in contrast to other elements of a piece of writing. Lexical means: pertaining to words or relating to words or the vocabulary of a language according to the Oxford dictionary. From the foregoing, it can be deduced that lexis is equivalent to the words of a language. Related terms are like lexicon, lexicography, lexeme and lexicology.

Singleton (2000), states that the term lexicon is a Greek term, which means 'dictionary' literally. It is used by linguists to refer to those aspects of a language, which relate to words otherwise known as the lexical aspects. Lexicon is based on

the term *lexis* which in Greek means word but which is used in linguistic terminology for vocabulary.

*Lexis* therefore, is the study of words or the vocabulary of a language. *Lexis* has been studied from various perspectives in recent years. Mburu (2003) describes and categorizes the lexical and phonological errors in the spoken English of pre-school level teachers in Nairobi. Mburu's study examines errors that emanate from inappropriate selection of lexical items but does not look at multiword units. Kellerman (1978) as quoted by Mburu (2003) researched on learner use of a set of idiomatic expressions and found that L2 learners avoid using idioms with peripheral meaning. Mburu's study also cites Levenston's work, which observes that L2 learners avoid selecting some lexical items, which would naturally be picked by a native speaker.

Nyamasyo (1992) investigates the lexical characteristics of the writing of Kenyan pre-university students. She uses corpora of the written English of 'O' and 'A' level students and highlights various lexical errors such as homonyms and synonyms errors.

Although Mburu (2003) and Nyamasyo (1992) concern themselves with lexical errors, their studies are relevant to this study by virtue of the fact that they deal with *lexis*. In addition, the present study seeks to analyze George Orwell's novel

*Animal Farm* for multiword units- with a view to making recommendations for pedagogic action. Hence, the views of the above-cited researches are relevant to the present work.

Nyamasyo (1992) also uses a written corpus and since, this research is corpus based, the approach she uses in extracting lexical items from the corpus could inform this study in the identification of multiword units in *Animal Farm*. Ndung'u (2002) has researched on lexical choices that learners make in line with the aspect of co-occurrence. Her research deals only with those patterns determined by synonymy. Since the present study deals with collocations, Ndung'u's research gives insights on the aspect.

Ongele (2005) examines the suitability of corpus instruction in the teaching of phrasal verbs. The fact that she uses a corpus makes her work relevant to this study. In addition, she deals with phrasal verbs, which is one type of the multiword units this study seeks to isolate from the novel *Animal Farm*.

### **2.1.2 Multiword Units.**

A multiword unit consists of two or more words whose grammatical form and lexical content, is wholly or largely fixed (Ongele, 2005). A major discovery of corpus linguistics has been lexical collocation as Sinclair (1987) notes. Broadly speaking, collocation includes groups or pairs of words with strong co-occurring

patterns within a limited amount of discourse (Gardner, 2007). Gardner goes on to say that, multiword units are collocations that are fairly fixed and are referred to as formulaic language. A multiword unit is a vocabulary unit, which consists of a sequence of two or more words. The sequence of words semantically and / or syntactically forms a meaningful and inseparable unit. Gardner argues that multiword units are not marginal phenomena since multiword units may collectively exceed the number of individual words in the lexicon. The units may also have multiple meanings considering that: the 104 most frequent phrasal verbs in the British National Corpus such as break up have 550 potential meaning senses, or an average of 5.6 per phrasal verb.

Sinclair, (2004) postulates the maximal approach where all the relevant patternings of a word are included. Consideration is also made of all patterning instigated by the central word. The unit is extended until all the ambiguity disappears. To Sinclair, a lexical item is one or more words that gather to make up a unit of meaning.

Singleton, (2000) says that collocation is the phenomenon of words keeping company together. It comes from two Latin words cum (with) and Locus (place). Words that form collocations are placed with each other; they often co-occur within a short distance of each other in spoken and written contexts such as novels. Multiword units are examples of stable collocations.

Collocation patterning has implication on our understanding as Singleton (2000) observes. He says that in order to function smoothly in lexical terms in any language, knowledge about collocation patterns is paramount. A listener would not be surprised if a news bulletin mentioned *breaking news* since the association between the two is so common that it has come to be regarded as a single term. Other collocations such as *ethnic cleansing* sound horrible yet the constituents are harmless when in isolation. On the same token, if an excited person said, *I am on top of the moon*; he/she would run the risk of incomprehension.

Schmitt (1997) defines multiword units as strings of characters or sequences of one or more morphemes, bounded at either side by a space or punctuation. The present study considers lexical connections between words in novels and makes recommendations for pedagogic action. Collocation and corpus studies are linked since it is difficult and pointless to study such things except through using large amounts of real time data.

Multiword units are distinguished from other strings based on the following criteria: Institutionalization, fixedness and non-compositionality (Schmitt, 1997). Institutionalization is the degree to which a multiword unit is conventionalized in a language. Fixedness is the degree to which a multiword unit is frozen as a sequence of words. For example, does it inflect? An item such as *rock the boat*

may not be restated as: '*rock the boated*', for example. Fixedness also considers if an item varies in its component lexis or word order.

Non-compositionality is the degree to which a multiword item can be interpreted on a word-by-word basis, but has a specialized unitary meaning. *Kicks the bucket*, for instance has no connection with a bucket. The above criteria are not absolute but variables, they work in conjunction with phonological criteria in speech where they form single tone units.

There are various categories of multiword units. These are compounds, phrasal verbs, idioms, fixed phrases and prefabs according to Schmitt (1997). Multiword units are therefore a super-ordinate term. Compounds are the largest and most tangible group- they differ from single words only by being written as two or more words. Many open or two compound words such as *crystal ball* and *commander-in-chief* are nouns. Compounds are mainly fixed but their non-compositionality and institutionalization varies.

Phrasal verbs are combinations of verbs and or adverbial or prepositional particles. The verbs are typically monosyllabic. It is a very prolific category especially for verbs like: come, go and put. The most common particles are: off, in, on and down. They have specialized meanings, which range in compositionality from transparent ones like: writing down, through completives such as *eating up* to

opaque ones like *tick off*. They are stylistically heterogeneous and are at times unmarked. *Give up* may mean cede, yield and relinquish. Some like *chill out* and *hang out* are jargonistic. Lastly, phrasal verbs are viewed as arbitrary unanalysable combinations that cannot be rationalized.

The other category according to Schmitt is the fixed phrases. These units fall outside other domains. They are like: *of course*, *at least*, *in fact* and *by far*; others are greetings and phatics such as *good morning* and *excuse me*. They are highly institutionalized with a high frequency of occurrence. Many are strongly fixed with variable non-compositionality; they have a specialist non-predictable function. Other fixed phrases are proverbs such as *it never rains but it pours*, which are included though infrequent.

Other types of multiword units according to Schmitt are idioms. These are multiword units, which are not the sum of their constituent parts. Their meanings are holistic and not retrievable from the individual meanings of their constituents a good example is- *spill the beans*. Moreover, they are historical, metaphorical and etymological with regard to a particular culture. The metaphor may be straightforward such as *bite off more than one can chew* or obscure such as *kick the bucket*. They are highly non-compositional, frozen in terms of fixedness, but not highly institutionalized.

The last category is prefabs: this is the most widely researched area in idiomatology. They are prefabricated routines or ready-made chunks according to Schmitt (1997) who calls them lexical phrases. Prefabs are preconstructed phrases, phraseological chunks and stereotyped collocations on semi fixed strings. They are tied to discursal items. They form structuring devices such as: *the thing is* and *that reminds me*. Prefabs are frozen in spite of varying widely.

Schmitt (1997) also observes that particular types of multiword units are preferred in particular genres. Idioms -for example, are mainly found in conversations, journalism and horoscopes while compounds are found in technical writings.

Carter (1998) calls multiword units fixed expressions. He says they present difficulties to non-native learners of language. This is because they cannot be understood literally, as some are restricted collocations like idioms. Carter also looks at irreversible binomials; full idioms; semi idioms; proverbs; stock phrases; catch phrases; allusions; idiomatic similes and discursal expressions.

Multiword items have several characteristics: these are varying collocation patterns and degrees of opacity. An examination of how words are placed together or co-occurrence reveals that, some words enter into a greater number and variety of lexical partnerships while others are 'choosy' about their combinations (Singleton, 2000). A word like *nice*, combines with a variety of words while

*addled*; which means literally rotten, connects only with the word egg. In its metaphorical, the word *addled* combines only with mind and brain. Collocation range is the term that defines various patterns of combination. *Nice* has a very wide collocation range while *addled* has a restricted range.

Opacity according to Carter (1998) is the relative fixedness of different multiword items. Here, Carter is in agreement with Singleton (2000) who states that multiword units are not evidently transparent or opaque but have varying degrees of opacity. Carter mentions the usefulness of clines. A cline being a scale that captures essential features of a linguistic phenomenon by seeing items as being distributed along continua of relative fixedness.

Some multiword units according to Carter are transparent semantically while others are fixed. No multiword item collocates with just any item - they are domain restricted. For example, *further to my letter* is found in business correspondence while *in no uncertain terms* is found only with performatives.

Schmitt (1997) underscores the importance of zeroing in on multiword units in language research. He fronts for the use of large corpora in collocation studies. Schmitt considers collocation to be a crucial principle that underlies the structure of language. Further, he says they are numerous and ought to be explored for their potential contribution to linguistics and language teaching. A corpus yields the

best results and the current research is carried out in line with the views of Schmitt (1997) since his categorization is clearly defined compared as opposed to that of Carter. Although the researcher adopts Schmitt's categorization, Carter's notion of a cline is used in the analysis of data on opacity and patterning from the novel.

There have been various researches on multiword units. Ongele (2005) examines phrasal verbs albeit in the teaching context. Howarth (2007) has researched on the relevance of phraseology to an understanding of non-native speaker academic writing. Howarth, stresses the central role of phraseology saying that: much of what people say is already stored in their lexicon in form of word combinations and manipulated in chunks. His research is relevant to this study as it concludes that the collocation complexion of a non- native text reflects the extent the writer possesses native – speaker competence. He also says that the non-use of restricted collocations leads, to imprecision. Howarth also states that lack of phraseological competence is a deficiency even in advanced users and refers to the works, which would assist an independent learner. Howarth criticizes the use of non-standard collocations, which hinder the comprehension of a non- native speaker's academic writing. Another point that he makes is that language researchers must redefine context and not just exhort learners to learn vocabulary 'in context' without telling them what context means. Howarth's view, are relevant to this work in that they help in the articulation of recommendations.

Conklin and Schmitt (2007) have also carried out research on formulaic sequences. Their concern is with how fast multiword units are processed compared to non-formulaic language. They have found that both native and non-native speakers of English process formulaic sequences faster than non-formulaic ones. They conclude that it is possible for non-native learners to enjoy the same type of processing advantage as the natives. This is relevant to this research in that, it shows how important formulaic sequences are; on the same note, the current study seeks to identify formulaic sequences from the novel *Animal Farm*. Recommendations for pedagogical action are then made based on the aspects uncovered on the multiword units.

Finally, Vintar (2007) has carried out a lexical analysis of a Slovene English parallel corpus. The conclusion the research draws is that multiword units are crucial to other branches of linguistics such as lexicography and translation and must therefore be treated properly. Spela's research helps the current one as it shows how multiword units are extracted from a corpus.

### **2.1.3 Language and Literature**

Various works have discussed linguistic aspects of literary works. These are like Leech (1969) who discusses how one could carry out a linguistic research on English poetry. Leech discusses the features of poetry such as sound features like rhyme. His work helps to justify the choice of a literary text in the present study.

In addition, Culler (1975) explains how the structuralism theory of linguistics may be used to analyze literary works-his rationale being that, literature is founded on language. The works of Culler and Crystal (1969) who writes on how one may investigate English style based on conversations and commentaries; written and spoken material also helps establish the connection between language and literature. Collie and Slater (1992) note that although; the literary world of novels and plays is a created one, it offers a full and vivid context in which characters from many social backgrounds are depicted. This offers an indirect route to understanding how language is used in the culture of origin.

## 2.2 Theoretical Framework

This research is informed by two theories: the Idiom Principle of Sinclair (1991) and the Lexical Approach (Lewis, 1993). Sinclair (1991) posited the Idiom Principle in his central paper on collocations. The theory views meaning as chunks of language that are more or less predictable. Alongside the idiom principle, Sinclair came up with the open choice principle. The two principles coexist and complement each other.

The open choice principle is a way of seeing a language text as the result of a very large number of complex choices. At each point where a unit is completed a word or phrase or a clause a large range of choices opens up, and the only restraint is grammaticalness (Schmitt, 1997). Hence, the open choice principle is a traditional approach to the study of language. It resembles the traditional slot and filler exercises approach, and the syntagmatic frame and paradigm distinction. This may also be expressed as the syntax governed structural frame and the lexical grammatical set of words respectively.

The Idiom Principle on the other hand states that a language user has available to him or her large numbers of semi pre-constructed phrases that constitute single choices, even though they might appear analyzable into segments (Schmitt, 1997). The Idiom Principle then restricts the choices not just in a given slot but also in the surrounding co-textual slots. Since this research investigates instances of

multiword units in *Animal Farm*, the two principles help in the identification of the multiword items; as well as in the analysis of the data so collected. Using the Idiom Principle, and the open choice principle, the research categorizes the multiword units into four classes embodied in the following examples from the novel: *kicked off* (phrasal verb) *cleared his throat* (idioms), *side by side* (prefab) and *hen houses* which is a compound word.

The second model adopted in this research is the Lexical Approach, which was propounded by Lewis (1993). Lewis states that, mental lexicon is not just a list of words - that is why he proposed his version of the lexical approach. He argues that: language consists of grammaticalised lexis not lexicalized grammar. He also refutes the traditional view that, language competence consists of having a foundation of grammatical structures into which individual words are slotted.

On the contrary, we store a huge arsenal of memorized phrases, words and collocations along with their associate grammars. To speak fluently, we draw from this vast phrase book, the chunks we need to converse effectively. Willis (1994) also echoed the views of Lewis when he said that language operates on the ability to deploy a wide range of lexical chunks both accurately and appropriately Willis (1994). This is what distinguishes advanced learners from intermediate ones. Both Lewis and Willis acknowledge the crucial function of vocabulary and question the traditional distinction of grammar and vocabulary.

They claim that words are small grammar, and grammar is big words. Their argument is that lexis ought to play a more prominent role in language planning, teaching and syllabus design. Lewis says that within the lexical approach, less attention is given to individual words and much less to traditional grammar structures, while the most attention is given to creating awareness on lexical items and collocations; that carry much of the meaning in spoken and written texts are examined for the kind of structures embedded in them.

The central tenets of the lexical approach according to Lewis are:

1. The grammar / vocabulary dichotomy is invalid
2. Collocation is used as an organizing principle.
3. Successful language is a wider concept than accurate language.
4. The observe – hypothesis – experiment replaces the present – practice – produce paradigm.
5. More importantly, language consists of grammaticalised lexis not lexicalised grammar.

To conclude, the lexical approach deals, mainly with language teaching techniques that are lexical in focus. Nonetheless, the approach helps in highlighting the multiword prefabricated chunks, in conjunction with the idiom principle. In addition, it helps in the data quantification since its claim that language is actually:

chunks or prefabricated phrases help account for the occurrence of multi-word units in the novel. The two theories are founded on collocation research, which is the concern of this study. This explains the choice of the two as the theoretical constructs.

## CHAPTER THREE

### RESEARCH METHODOLOGY

This chapter discusses the methodology in terms of: research design, the study population, the sampling design, and the data collection and analysis procedures.

#### 3.1 Research Design

This study is library based mainly due to its textual nature. It entails a detailed investigation of multiword units in George Orwell's novel *Animal farm* to find out specific things about them. Consequently, the research adopts a case study design.

The research describes the multiword units in detail, contextually and holistically. It uncovers specific things about multiword units and a great deal is learnt from a few examples of the items that have been studied. A case study best suits the research, as it helps the research to intensively focus on the novel and identify the various multiword units with ease.

#### 3.2 Sampling Techniques and Sample Size

The researcher focuses on the whole text. This means that, there was no sampling as the novel is short enough to be read in one sitting. In addition, multiword units only comprise a third of both written and spoken language. Consequently, the researcher handles all the multiword units in the eighty five-page novel. Another

justification for not sampling the text is that: unless the whole text was analyzed, the authenticity of the data as well as the accuracy of the frequencies would be questionable. It was therefore necessary to draw data from the entire text which also ensures the data is representative. Finally, the target population that is: all the multiword units in the novel are small such that, selecting a sample would be meaningless.

However, at the point where the data is analyzed by use of clines, the research randomly selects ten multiword units in each of the five categories for analyses. This helps the researcher select the units that are most informative with regard to the second and third objectives. Again, this prevents a duplication of findings since a few units are sufficient to show the aspects of opacity and patterning of the multiword units in the novel. In any case, all the multiword units have been analyzed statistically at objective one level hence; any sampling conducted later on is justifiable and representative of the entire population. Furthermore, any other number of multiword units would fit within the specified clines. In fact studies by Milroy (1987) and Sankoff (1980) have shown that one does not require a very large linguistic sample to observe a linguistic phenomenon, the only restriction being representativeness. The decision not to sample at the data mining stage and the first objective level is purely to ensure that the frequencies and charts in objective one are accurate and authentic.

### **3.3 Data Collection**

The data required for this study is lexical items. There are single word lexical items and multiword lexical items. The present study is primarily concerned with multiword lexical items. Since this research is textual in nature, the multiword units are identified and extracted through content analysis. The units are extracted manually although there are machine extraction programs. The reason is that: for one to obtain data from a text, it need not be machine-readable. The idioms principle and the lexical approach theories enable the researcher to single out the multiword units from other lexical items in the text.

### **3.4 Data Analysis**

The data so collected is categorized into: compound words, phrasal verbs, idioms, and prefabs in line with the categorization of Schmitt (1997). The units are then keyed into a computer. Thereafter, qualitative and quantitative procedures are used to analyze the data. The quantitative techniques adopted are percentages, graphs and charts. This helps account for the total number of instances for a given category of multiword units.

The qualitative technique adopted is content analysis. Themes and patterns are identified for select multiword units. The technique is used to describe the degree of opacity as well as the patterns of collocation inherent in the multiword units. The research then considers the semantic opacity; by using the notion of a cline to

range the multiword units from the least to the most opaque- in terms of sets of continua with fixed points but several intermediate categories. The degree of opacity is described using the following table.

**Table3.1: The semantic opacity cline for various multiword units**

TRANSPARENT	SEMI-IDIOMS	SEMI- TRANSPARENT	OPAQUE

The collocation patterns are analyzed by introducing the notion of a cline. The multiword units are graded in terms of restriction, where the cline runs from the less fixed to the most fixed. This implies the use of a continuum with fixed points but several intermediate categories. The continuum comprises of: unrestricted collocations, semi- restricted collocations, familiar collocations and restricted collocations. The table below captures this assertion pretty well.

**Table3.2: The restriction cline for the patterns of various multiword units.**

UNRESTRICTED COLLOCATIONS	SEMI-RESTRICTED COLLOCATIONS	FAMILIAR COLLOCATIONS	RESTRICTED COLLOCATIONS

As earlier mentioned, the researcher does not use sampling in objective one but in the second and third objectives; sampling is done since the researcher is observing the opacity and patterning of multiword units which can be illustrated using a small sample as posited by Milroy (1987) and Sankoff (1980). Ten multiword units in each of the four categories: phrasal verbs, compound words, prefabs and idioms are purposively selected and analyzed for semantic opacity and collocation patterns- using clines.

## CHAPTER FOUR

### DATA ANALYSIS AND PRESENTATION

#### 4.0 Introduction

This chapter consists of the presentation and analysis of the multiword units in George Orwell's *Animal Farm*. The data was organized with respect to the research objectives so as to generate an easy – to - follow discussion (Paton, 1990). This study seeks to investigate the types of multiword units, the opacity of multiword units and the patterns of collocation in multiword units with illustrations from *Animal Farm*. There are two stages in the data analysis. To begin with, the multiword units are identified and categorized into: compound words, phrasal verbs, idioms and prefabs as advocated by Schmitt (1997). Although the research had proposed to consider five categories of multiword units as discussed in Schmitt (1997), it is found that there are no constituents that can be classified as fixed phrases so that, only compounds, phrasal verbs, idioms and prefabs are identified. A table showing the number of multiword units in each of the five categories is then drawn after which percentages are generated from the results. The information is then presented in the form of a pie chart so that the distribution of multiword units in the data can be highlighted more clearly.

The second stage involves purposively sampling ten multiword units; from each of the four categories of multiword units determined in stage one of the data analyses.

The units are then subjected to restriction and opacity clines. A cline refers to changing degrees of an aspect when measured against a scale (Carter, 1998). Sampling is done in order to get a representative sample. Again, one does not need to have a very large sample so as to describe a linguistic phenomenon (Milroy, 1987)

The restriction cline runs from the unrestricted to the restricted collocations with semi-restricted and familiar collocations as the intermediate categories. The opacity cline runs from the transparent to the opaque collocations with semi-idioms and semi-transparent collocations as the intermediate categories. The analyses first consider the restrictedness of the multiword units and then examine the opacity of the multiword units.

#### **4.1 The Multiword Units in the Data**

The analyses identify a total of 447 multiword units from the text. In total, there are 121 compound words, 98 phrasal verbs, 63 idioms and 165 prefabs. However, no fixed phrases are identified. A complete list of the multiword units may be viewed in appendices 1,2,3 and four.

#### 4.1.1 Compound Words

A compound word is a combination of two or more lexical items which form a new lexical unit with a meaning that is totally unrelated to that of the constituents. For example, *bedroom*, *birch saplings* and *bodyguard* (Appendix 3). The compound words are identified and classified on the premise that: a compound word differs from a single word by being written as two or more orthographic words (Schmitt, 1997).

The analyses find that compound words can be divided into three categories on the basis of their orthography. That is to say, some are fused with their constituents combining completely as in *mantelpiece*. Others are hyphenated which means the constituents are separated by means of a hyphen; consider the compound adjective *sly-looking*. The third sub-category of compound words does not comprise of open compound words whose constituents are neither fused nor do they have a hyphen. The word *pop holes* is a case in point. Table 4.1 shows the proportions of fused, open and hyphenated compound words in the data.

**Table 4.1 The structural grouping of compound words**

Class	Fused compound	Open compounds	Hyphenated compounds	Total
Nouns	41	34	27	102
Verbs	9	0	0	3
Adverbs	2	1	0	3
Adjectives	5	0	3	9
Total	57	35	30	122

As table 4.1 shows, the data has 56 fused compounds, 35 open compounds and 30 hyphenated compounds across different word classes.

#### 4.1.2 Phrasal Verbs

A phrasal verb is a combination of a verb and a prepositional or adverbial particle such as *keep off* – whose particle is *off*- a preposition while, *keep* is its verb stem. Phrasal verbs in the data are categorized as multiword units since they are lexical strands that have more than one word whose resultant meaning is non-compositional

The data identifies 98 phrasal verbs, *adapted to* and *broken of* being some of the phrasal verbs. The other phrasal verbs are in Appendix 4. The verbs in the phrasal verbs are mainly monosyllabic an example being *come into* but there are disyllabic

ones too; a case in point being, *galloped away*. The phrasal verbs are also described on the basis of whether they are separable or not. *Crept out* is an example of an inseparable phrasal verb from the data since its elements are always placed together in a given context.

*Keep off*, on the other hand is a separable phrasal verb since its constituents can be separated by an intervening word(s). This is well illustrated in the following examples.

(a) God had given him a tail to *keep the flies off*.

In the above example, the verb (keep) and its particle (off) are separated by two words – *the* and *flies* hence, *keep off* is a separable phrasal verb.

(b) While Major was speaking, four large rats had *crept out* of their holes.

In the example above, *crept out* is a phrasal verb. It is inseparable because the stem and the particle have to be placed together if the sentence is to make sense. For instance, if one said “Four large rats had *crept* of their *out* holes” the resultant sentence would be meaningless. This contrasts sharply with *keep off* since we could say “God had given him a tail to *keep off* the flies” where the sentence makes sense whether the particles co- occur or are separated from the verb.

The data has 40 inseparable phrasal verbs and 48 separable ones. *Come to* and *privy to* are examples of inseparable phrasal verbs; while *pass on* is a separable phrasal verb. Both separable and inseparable phrasal verbs can function transitively or intransitively. The analysis then concludes that phrasal verbs behave just like ordinary verbs due to the fact that they all have verb stems; the stems cause them to inflect for tense and number for example, just like ordinary verbs. In Appendix 4, the inseparable phrasal verbs are written in bold while the separable ones are written in normal print.

#### 4.1.3 Idioms

Idioms are multiword items whose meanings are not the sum of the meanings of their constituent parts: they have holistic meanings which cannot be retrieved from the individual meanings of the component words, (Schmitt, 1997). In lexical studies the term idiom may have a variety of meanings but in the present study, it is used in a narrow sense to describe the multiword units with meanings that do not equal the sum of that of their constituents. Examples of idioms from the data are; *driving hard bargains* and *to fall an evil days*. This research identifies 63 idioms, which are then divided into two categories: institutionalized and non – institutionalized idioms. Institutionalized idioms are those with a high frequency of usage in English while the non-institutionalized idioms are the ones that are rarely used. There are 27 non- institutionalized idioms and 36 institutionalized ones in the data. *Crystal clear* is an example of an institutionalized idiom while *to turn*

*black into white* is an example of a non-institutionalized idiom. This shows that most of the idioms in the data are institutionalized in English. A complete list of idioms in the data is available in Appendix 5. Non-institutionalized idioms are in bold while the institutionalized ones are in normal print in Appendix 5.

#### **4.1.4 Prefabs**

Prefabs are pre-constructed phrases, phraseological chunks, stereotyped collocations or semi-fixed strings which are tied to discursal situations and which form structuring devices (Schmitt, 1997): (Cater, 1998). The data analyses 165 prefabs examples being: *round and round*, *by heart* and *side by side*. The full list of prefabs is found in Appendix 6. The prefabs are all institutionalized in the English language and have, therefore, not divided into sub-categories; this means that they are strands that are used routinely by speakers of the language, which accounts for their high incidence of occurrence in the data.

#### **4.1.6 Conclusion.**

In conclusion, the above categorization is used in identifying and drawing attention to the range of multiword units in the data. Inevitably, there are overlaps between the categories hence; some of the elements can fit into more than one category. The term prefabs and fixed phrases as discussed in (Schmitt, 1997) is inappropriate in the analyses as a clear distinction cannot be made between the two categories. Therefore, the research classifies all the items outside the categories of

phrasal verbs, compounds and idioms as prefabs. It, therefore, follows that there are no hard and fast rules in the categorization of the multiword units but for the sake of convenience the multiword units are grouped simply to ensure the analyses of the data are consistent. Other overlaps are observed between idioms and phrasal verbs. *Give up* is an idiom in the true sense of the word but due to its lexical composition is termed as a phrasal verb too. All in all, the data shows that multiword units are very frequent in the data but much overlap ensues between various categories.

**Table 4.2 Distribution of multiword units**

Multiword units	Number of units	% Of total number
Compound words	122	27%
Phrasal verbs	98	23%
Idioms	63	14%
Prefabs	164	36%
<b>Grand total</b>	<b>447</b>	<b>100%</b>

Table 4.2 shows that prefabs far out number all the other types of multiword units in the data. Compound words come second followed by phrasal verbs while idioms are the least frequent. The research attributes the high frequency of prefabs to their being expressions that have been routinised in English. Much of language comprises of prefabs according to Sinclair (1987). Idioms are the fewest due to the

fact that they are the least institutionalized category in language. The same information is summarized in the pie chart below

**Fig.4.1 Pie chart showing the frequency of multiword units**

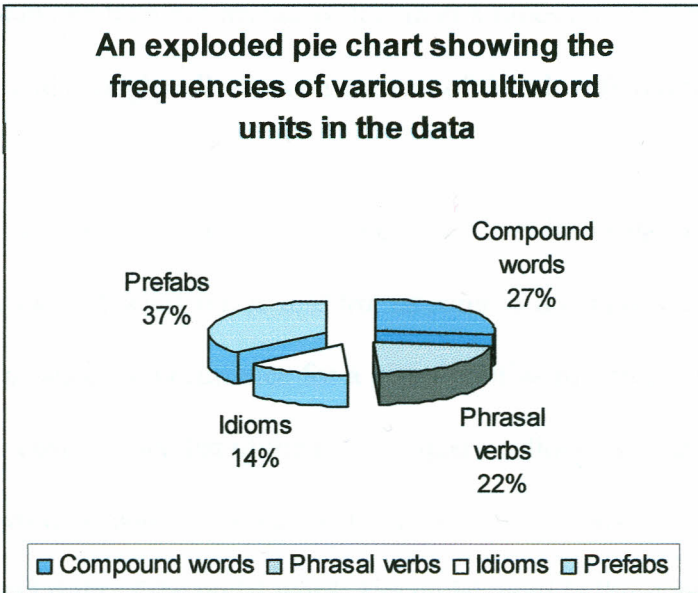


Fig 4.1 presents the distribution of multiword units in the text albeit differently from table 4.1. It also shows that indeed the most frequent items in the data are prefabs while idioms are the least frequent multiword units in the data.

## 4.2 The Restriction and Opacity Clines

In this section, ten multiword units in each of the four categories are purposively sampled and subjected to the restriction and opacity clines. A cline is a scale that ranges data along a continuum with two extreme points and various intermediate categories (Carter, 1998). To start with, the frequency of occurrence for each of

the sampled multiword unit is determined and those found to occur less than six times are left out as they cannot yield sufficient data. Consequently, only the units with an occurrence of 6 or more times are subjected to the restriction cline. Out of 40 sampled multiword units only 8 have an occurrence of six or more times. The units with an occurrence of less than 8 times are 32. These multiword units cannot yield enough collocates on account of their low frequency in the data.

The eight items analyzed in the data are: *side to side*, *round and round*, *farmhouse*, *cowshed*, *windmill*, *to and fro*, *as soon as* and *nevertheless*. In the analyses all the instances of occurrence for a given multiword unit are identified so as to establish a concordance list of the most frequent collocates. The multiword unit is the nodal word around which collocates cluster. The analyses consider 5 words to the left and right of the nodal word. The nodal word in this case is the multiword unit. The types or word classes with which the node collocates are also considered so as to shed more light on the patterns. The most frequent collocates in each case are the ones that have been highlighted in bold. Collocates that have been discussed in the data are lexical words that co-occur with the node while functional words are ignored.

Concordance lines with the multiword units as the nodal words are established after which the principal collocates are identified along with their positions before or after the node. A maximum of five words to the right and left of the node are

considered in the determination of collocates for a given multiword unit. All collocates of a given item are identified and thereafter, the most frequent collocate is singled out. Restrictedness is then established by considering the strength of association between the multiword units and collocates. If the association is strong, the multiword unit is classified as restricted but if the association is weak, the multiword unit is categorized as being unrestricted.

Restriction is the readiness with which a multiword unit partners with other lexical units. In the present analyses, multiword units are divided into four categories depending on the degree of restriction (Carter, 1998). There are unrestricted collocations in which the node partners freely with other lexical items. Secondly, there are semi-restricted collocations in which the range of items the node can partner with is more determined than in the unrestricted category. Familiar collocations on the other hand are combinations in which there is a regular association between the nodes and collocates. Finally, there are restricted collocations in which the partnerships of the node are fixed and closed; they do not permit substitution.

Opacity is the degree to which the meaning of multiword expressions can or cannot be seen as a straightforward composite of its component words (Singleton, 2000). Multiword units are not self-evidently opaque or self-evidently transparent but have various degrees of opacity. It is, therefore, against this background that

the multiword units are sub-divided into transparent, semi –idioms, semi-transparent and the opaque categories. The degree of opacity is found to be dependent on the context of usage so that the context of a multiword unit also has to be examined. It should be noted that the term cline presupposes a continuum in which case overlaps are definitely observed. In fact it is just a question of what multiword units have a higher degree of opacity compared to the members of the preceding category. There are: transparent multiword units in which one can “see through” the constituents, semi-idioms, which are metaphorical in nature, semi-transparent multiword units which are opaque but not fully and finally, opaque multiword unit which are fully opaque in the sense that a knowledge of the native English culture or prior encounter with the multiword unit is a prerequisite to the correct interpretation of the meaning. It must be emphasized from the outset that no multiword unit is completely restricted or opaque but what the research sought to do was to determine the extent to which a given multiword unit manifests restriction and opacity vis-à-vis the others in a given cline.

#### **4.2.1 Collocation Restriction in the Multiword Units**

Collocation is the phenomenon where words keep regular company. The said words tend to associate very much with certain lexical units. In the data for example, the phrasal verb *keep off*, associates closely with the nouns *tail* and *flies*. Thus, *tail* and *flies* are collocates of the phrasal verb *keep off*. Restriction refers to the degree to which a multiword unit partners freely with other lexical items.

Carter (1998) describes how the restriction of multiword units may be described by means of a cline. A cline is a continuum that determines the degree of an aspect exhibited by a linguistic element when measured against a predetermined scale. The scale has two extreme points and intermediate categories in between. The analyses classify the sampled multiword units into: unrestricted collocations, semi-restricted collocations, familiar collocations and restricted collocations. Unrestricted collocations have a wide collocation range which means they partner with a large number of lexical items. Semi-restricted collocations on the other hand embrace a number of items which can be substituted into more determined syntactic slots. Familiar collocations comprise of elements that keep regular company with each other. The fourth category is that of restricted collocations which comprises of multiword units that have the narrowest collocation range- this implies they are only open to partnership with a small number of items (Carter, 1998). The restriction cline according to Carter runs from the least restricted to the most restricted lexical units.

#### **4.2.2 Unrestricted Collocations**

In the data the unrestricted collocations are: *farmhouse*, *windmill* and *nevertheless*. These multiword units have a wide collocation range as Singleton (2000) observes. The following examples illustrate the concept of unrestrictedness. The collocates of the compound word *farmhouse* are: *door*, *plans*, *moved*, *animals*, *found*, *pigs*, *suddenly*, *beds* and the other five words to the left and right of the nodal word.

The most prominent collocate is the noun *kitchen* which occurs twice in the data. It succeeds the word *farmhouse* in the two instances of its occurrence. The word *farmhouse* is unrestricted since it partners with a broad range of lexical units as the highlighted examples in bold show. With the exception of the word *kitchen*, *farmhouse* does not have a preferred collocate simply put, it partners with very many words in the data. Below are the instances in which the word is used in the data. It mainly partners with verbs such as *found* and *was* as well as nouns like *kitchen*.

#### Example 1: farmhouse

-to the farm buildings and halted in silence outside the door of the **FARMHOUSE**. That was theirs too, but they were frightened...

-which he had found in the **FARMHOUSE**, and was full of plans for innovation...

3-that the pigs suddenly moved into the **FARMHOUSE** and took up residence there.

‘That we pigs sleep on the beds in the **FARMHOUSE**? And why not? You did not suppose....

-piled on china dish from the **FARMHOUSE** kitchen. The animals filed slowly past, and...

-a school room would be built in the **FARMHOUSE** garden. For the time being..

-young pigs given their instruction by Napoleon himself in the in the **FARMHOUSE** kitchen. They took their exercise in the garden...

The other unrestricted word in the data is *nevertheless*. This compound word combines with many words and does not have particular words with which it partners. The example below shows that the principal collocate of the word *nevertheless* is **Napoleon** which occurs twice. Some of the other collocates are *animal farm*, *towards*, *shortage* and *interests*. The five words to the left and right of the node are the ones that partner with *nevertheless*; to the left and to the right of the node in the data. The concordances show that the word *nevertheless* is unrestricted in its partnerships as far as the data is concerned. Another point is that collocates are mainly verbs and nouns such as *laugh* and *farm* respectively.

#### Example 2: nevertheless

-even in the defense of their own interests. **NEVERTHELESS**, they were both thoroughly frightened by the...

-that there was no food shortage on Animal Farm. **NEVERTHELESS**, towards the end of January, it became...

-and trust in comrade Napoleon's strategy. **NEVERTHELESS**, feelings against Fredrick continued to run high.

-say that he saw nothing to laugh at **NEVERTHELESS**, without openly admitting it, he was devoted to Boxer.

- and avoided him as much as possible. **NEVERTHELESS**, the sight of Napoleon on all fours, delivering orders to whimper...

-a lot of labour on the upkeep of hedges and gates. **NEVERTHELESS**, as the summer wore on, various unforeseen shortages...

Windmill is an unrestricted collocation since it partners with very many words freely; as the words to the left and right of the node show. In the data, *windmill* combines with words such as *just*, *place*, *starve* and *death*. The word windmill does not have a most frequent collocater as the examples below show. For that reason it is classified as unrestricted. Collocates of the word windmill occur once or twice and these collocates are mainly nouns and verbs such as death and starve respectively.

### Example 3-Windmill

Snowball declared that this was just the place for a **WINDMILL** which could be made to operate a dynamo

-and that if they wasted time on the **WINDMILL** they would starve to death

-he had seemed to oppose the **WINDMILL** simply as a maneuver to get rid of Snowball

Within a few weeks, Snowball's plans for the **WINDMILL** were fully worked out. The mechanical details...

He had declared himself against the **WINDMILL** from the start. One day however, he arrived unexpectedly to examine....

The whole farm was deeply divided on the subject of the **WINDMILL**. Snowball did not deny that to build it...

-their own special contribution towards the building of the **WINDMILL**. Once again, the animals were conscious...

#### 4.2.3 Semi-restricted Collocations

These are the collocations whose patterns have members that may be substituted into more determined syntactic slots. The only example in the data is *as soon as* which is classified as a semi-restricted collocation because it partners with lexical units which are found only in some syntactic slots. Examples of collocates of *as soon as* are: *Mr. Jones*, *Snowball*, *sturdy puppies and gun* which function as subjects and objects in various syntactic slots. For instance, when we are told that, “*Snowball had intended...*” *Snowball* in this case functions as a subject in that construction *flight* acts as an object in the sentence “*She had taken to flight as soon as...*” *As soon as*; is semi-restricted when compared to an example such as *farmhouse*, which does not seem to have any restrictions in its partnering. The five words to the right and left of the node *as soon as* are collocates. Although the word *as soon as* does not have a principal collocate, it is semi-restricted in the sense that it collocates with words that perform particular syntactic functions in the data as discussed above. Otherwise it collocates mainly with verbs and nouns just like in the previous examples of multiword units. *Sturdy-* *puppies* is a noun while *weaned* is a verb for instance. Lastly, it should be noted that the multiword unit *as soon as*

also performs a conjunctive role in the data which again shows the limitations placed upon it hence its categorization as semi- restricted. In the seven instances of its occurrence it makes the syntactic structures it is found among more cohesive.

#### Example 4-as soon

...them to nine sturdy puppies. **AS SOON AS** they were weaned, Napoleon took them....

This was just what Snowball had intended. **AS SOON AS** they were well inside the yard...

She had taken to flight **AS SOON AS** the gun went off. And when the others....

, and escorted him back to the farmhouse **AS SOON AS** he was able to walk.

None of...

Which had to be closely covered with writings, and **AS SOON AS** they were so covered, they were burned in the farmhouse.

They should all meet in the old barn **AS SOON AS** Mr. Jones was safely out of the way

- where Mrs. Jones was *already* snoring. **AS SOON AS** the light in the bedroom went out.

#### 4.2.4 Familiar Collocations

These multiword units associate regularly with collocates. Examples from the data are: *side to side*, *round and round* and *to and fro*. *Round and round*, *side to*

*side* and *to and fro* are prefabs. They are categorized as familiar collocations since they are observed to keep regular company with some of collocates. *Round and round* has regular collocates such as these verbs of movement: *chasing, trampled, gamboled, hurled, and walked*. These words are all verbs of motion and they all collocate with the prefab *round and round* hence the concept of familiarity. The noun *animal* is the most frequent collocate of the prefab *round and round*– it occurs thrice. The second most frequent collocate is the verb *gamboled* which occurs twice in the data. The other collocates are the five words to the left and right of the node in the following examples.

#### Example 5: round and round

–all the animals were chasing them **ROUND AND ROUND** the yard they were gored kicked, bitterly, trampled on.

– (no bits and reins were needed in these days, of course) and tramp steadily **ROUND AND ROUND** the field with a pig walking behind and calling out....

–an especially devoted follower of Napoleon by chasing him **ROUND AND ROUND** a bonfire when he was suffering from a cough

–in the ecstasy of that thought, they gamboled **ROUND AND ROUND**, they hurled themselves into the air in great leaps of excitement

–tired out but proud, the animals walked **ROUND AND ROUND** their masterpiece which appeared even more beautiful. In....

–in their spare moments, the animals would walk **ROUND AND ROUND** the half-finished mill admiring the strength....

–their tiredness forsook them and they gamboled **ROUND AND ROUND** the windmill, uttering cries of triumph .Napoleon himself....

*Side to side* is a familiar collocation because it collocates regularly with words such as *skipping*, *whisking*, *twitched* and *tail*. There is some kind of association between the node and those collocates which is strong. The most frequent collocate of the prefab *side to side* is *skipping*, which occurs thrice followed by *whisking* which occurs twice. *Skipping* precedes the prefab in every instance of occurrence in the data. The other collocates were like *pleading*, *rigid* and *sharply*. Another point is that most collocates of the prefab *side to side* are verbs of movement such as *whisking*. In addition, these verbs are dynamic in nature since the actions they represent are intense. The prefab *side to side* also has adjectives such as *sharply* and nouns like *tail* as collocates.

Example 6: side to side

–he had a way of skipping from **SIDE TO SIDE** and whisking his tail

–his tail had grown rigid and twitched sharply from side to side, a sign in him of intense mental activity.

–cried Squealer almost pleadingly skipping from **SIDE TO SIDE** and whisking his tail, ‘surely, there is no one...

-whisking his tail and skipping from **SIDE TO SIDE**, surely, they knew their beloved leader Comrade Napoleon....

.. majestically upright casting haughty glances from **SIDE TO SIDE** and with, and with his dogs gamboling round him.

-cheeping feebly and wandering from **SIDE TO SIDE** to find some place where they would

—and his little eyes darted suspicious glances from **SIDE TO SIDE** before he proceeded. It had come to his knowledge...

—with the ring of light from his lantern dancing from **SIDE TO SIDE**, he lurched across the yard, kicked off his boots the back door...

To and fro, is the only irreversible binomial identified in the analyses. The prefab associates a lot with verbs such as *toiled*, *fidgeted*, *swishing flew*, *plod* and *paced* which are verbs of movement. This is what makes the prefab a familiar collocation. Some of the other collocates are *sun*, *rapidly* and *carrying*. The prefab does not have a most frequent collocation. However, many collocates are verbs of movement which is what makes it a familiar collocation seeing most collocates relate to movement.

Example 7: to and fro

-even the ducks and hens toiled *to and fro* all day in the sun, carrying tiny wisps...

–Only Boxer remained on his feet. He fidgeted *to and fro* swishing his long black tail against...

–...gripped between the knuckles of his trotter, he would move rapidly *to and fro*, drawing in line after line and...

–all the pigeons to the number of thirty-five flew *to and fro* over the men's heads and dropped...

–thinking it well worthwhile to plod *to and fro* the all day with blocks of stone. Napoleon paced...

–at the sight of the fallen stone Napoleon paced *to and fro* in silence ,occasionally sniffing at the ground..

#### 4.2.5 Restricted Collocations

The only restricted multiword unit in the data is *cowshed*. This node only partners with a few elements in the data and rarely allows a combination with words outside the select group of collocates. Again, it should be emphasized that the item is termed restricted since compared to the other multiword units in the data, its node has a large number of words with which it partners. Hence, restriction is a question of the extent of the aspect in a set of multiword units.

Cowshed is the most restricted multiword unit in the data in that in all its instances of occurrence, it co-occurs with the words *in the battle of the* in the data. In almost all the occurrences of the word *cowshed* in the data, one encounters the

lexical unit *battle*. The other regular companions of the word *cowshed* in the data are *defeated* and *ambush* which also relate to the meaning of the word *battle*. Therefore, the compound word *cowshed* is considered restricted as it has the narrowest collocation range in the data. The most frequent collocates of the word *cowshed* is *battle*. The other collocates are the five words to the left and right of the node. *Cowshed* also collocates mainly with verbs and nouns an example of a noun being *battle* while that of the verb is *defeated*. The principal collocate of *cowshed* is *battle* which is a noun.

#### Example 8: cowshed

–‘Loyalty and bravely are more important. And as to the battle of the **COWSHED**, I believe the time will come when....

-he sniffed in every corner, in the barn, in the **COWSHED**, in the farmhouse in the hen houses, in the...

–‘I believe that at the battle of the **COWSHED**, he was a good comrade.’ Our leader Comrade Napoleon.’ announced Squealer...

It was fully realized that though the human beings had been defeated in the battle of the **COWSHED**, they might make another attempt and..

–What the battle should be called. In the end, it was named the battle of the **COWSHED** since that was where the ambush had been sprung...

–‘I do not believe that.’ he said. Snowball, fought bravely in the battle of the **COWSHED** I saw him myself. Did we not give him...?

–the pigs, who had who had been lying in ambush in the **COWSHED** suddenly emerged from their rear...

‘He attempted –fortunately without success, to get as defeated and destroyed at the battle of the **COWSHED?**’ The animals were stupefied. This was.....

### 4.3 Opacity in the Multiword Units

Opacity refers to the extent to which a multi word unit is not interpretable without contextual / cultural knowledge. Cruse (1988), emphasizes that transparency and opacity are end points in the continuum of the degree of opacity. Consequently, there are only ‘fully transparent’ and to some degree ‘opaque’ lexical items. Carter (1998) describes an opacity cline in which a continuum accommodates multiword units by subjecting them to a scale of semantic opacity. The cline has two extreme points (opaque and transparent points) and two intermediate categories, which are semi-idioms and semi transparent multiword units. Howarth (2000) asserts that transparent collocations have straightforward meanings, which may be derived from the original meanings of the constituents. These units are also used in their primary literal senses. Semi-idioms are also called metaphors or idiomatic similes. In these collocations, the constituents are used in a figurative or specialized sense only found in the context of the other element or a limited number of alternatives.

The other collocations are the semi – transparent ones. These multiword units have their composites being employed in a figurative sense as a whole. However, there is a literal interpretation, which is rarely used. Finally, there are opaque collocations in which the meaning of the unit cannot be derived from the sum of the meanings of the constituents. The item does not have a literal counterpart since the original habit or meaning of the multiword has disappeared and has been replaced by the present meaning of the item. The above categories merge into each other diachronically and synchronically. The multiword units that are subjected to the opacity cline were 18 as opposed to the ten that were subjected to the restriction cline. They are: *round and round*, *side to side*, *cowshed*, *set aside*, *farmhouse*, *windmill*, *cruel knife*, *five-barred gate*, *pass on*, *by heart*, *to and fro*, *nevertheless*, *drawing room*, *took to her heels*, *turn black into white*, *fallen on evil days* and *windfalls*.

#### 4.3.1 Transparent Multiword Units

The multiword units that are used in their literal sense are like: *round and round*, *farmhouse*, *cowshed*, and *side to side*. A few examples from the data will be used to illustrate transparency in multiword units. The multiword units mentioned above have all been used in a literal sense. In other words, their meanings are straightforward and recoverable from the individual meanings of their constituents. *Round and round* (see example 5) has a literal interpretation that is directly implied by the constituents which means circular motion. When it is said in the data that

‘the animals were chasing them round and round the yard’ it is possible to tell what is happening just by considering the individual meanings of the constituents. The same goes for *farmhouse* (see example 1), which means exactly that: the house on the farm. The examples from the data such as ‘*pigs sleep on the beds in the farmhouse*’ and ‘*moved into the farmhouse*’ all reflect the meaning of the compound word *farmhouse* in a straightforward manner hence the item is semantically transparent. *Side to side* in ‘he had a way of skipping from *side to side* and whisking his tail’ (also see example 6) is also semantically transparent since the meanings of the constituents mirror the meaning of the resultant multiword unit, which indicates some kind of movement that is directional and well coordinated. Finally; the compound word *cowshed* also has a straightforward meaning, which is directly recoverable from its constituents. In the expression ‘In the battle of the cowshed’ for example, it can be deduced that the venue of the battle is the place which shelters cows that is; the cowshed!

#### 4.3.2 Semi-idioms

These collocations are metaphors or idiomatic similes. One or both of their constituents are used in a figurative sense. Examples from the data are: *windmill/cruel knife/pass on/five-barred gate*. The word *windmill* (example 3) has a literal dimension in that it refers to a mill operated by wind power. However, it also has a metaphorical sense in that it is a power generation machine. It is, therefore, not possible to deduce the meaning of the multiword unit without

considering the context of usage as well as cultural knowledge upon which all languages rest. In sentence 1, example number three the meaning of the compound word *windmill* is semi-idiomatic. This is because without considering collocates of the word such as *could, be, made, operate* and *dynamo* the meaning can not be recovered easily. In view of the above, the word *windmill* is a semi-idiom. It does not mean a mill that produces wind but on the contrary, it refers to a kind of generator that utilizes wind power.

*Cruel knife* is another semi-idiom in the data as it has an element that is used in a figurative sense only in the context of the other. *Cruel* is the element that is used figuratively in the context of the word *knife* as in ‘no animal escapes the *cruel knife* in the end’ and the *cruel knife* with which Mr. Jones...’Now, a knife can never be cruel and the figurative usage of *cruel* to describe the knife makes the idiom semi-idiomatic in line with (Carter, 1998’s) claim. The examples are listed below.

Example 9: cruel knife

-no animal escapes the **CRUEL KNIFE** in the end

-the **CRUEL KNIFE** with which Mr. Jones

*Pass on* is a unique example in that its usage in example 10- sentence number 1 below is straightforward hence; the item is semantically transparent. However, in

sentence five, passed on in ‘to have *passed on* before the windmill was finished is semi-idiomatic as the original meaning of the constituents which means to convey information has been replaced by a new meaning namely, to die. Hence; some cultural aspects have to be factored in so as to deduce the meaning of the multiword, which means to die. The conclusion made is that some multiword units are either transparent or otherwise depending on the usage so that one has to go beyond the constituents to correctly deduce their meanings.

Example 10: pass on

-and above all, **PASS ON** this message of mine to the

-feel it my duty to **PASS ON** to you such wisdom as

-the rebellion was only a dim tradition, **PASSED ON** by word of mouth

-everything they were told and **PASSED ON** to the other animals

-his sole sorrow was to have **PASSED ON** before the windmill was finished

Five-barred gate in example 11 below is a multiword unit that is semi-idiomatic because it's meaning is not the sum of that of its constituents. On the contrary, it refers to a special kind of gate. It, therefore, follows that the term gate is literal while five-barred is metaphorical when used together with gate. In the five instances of the usage of five- barred gate below, we only get to know that it refers to the gate by considering the collocates of the word as in sentence four where we

are told 'had entered the five barred gate' the collocate enter sheds light as to the possible meaning of the word.

Example 11: five-barred gate

-His men out on the road and slammed the **FIVE-BARRED GATE** behind them and so almost before they knew what...

-his followers had already come through the **FIVE BARRED GATE**. Boldly enough, the animals...

-for pots of black and white paint and led the way down to the **FIVE-BARRED GATE** that gave on to the main road

-others from Pinchfield and Foxwood had entered the **FIVE-BARRED GATE** and were coming up the.....

-someone thought of racing ahead and shutting the **FIVE-BARRED GATE**; but in another moment, the van was through...

#### 4.3.4 Semi-transparent Collocations

The collocations with composites that are used as a whole figuratively i.e. all their constituents have a meaning that has to be inferred and they cannot be interpreted simply at face value. These semi-transparent collocations include *to and fro* and *by heart*

In example 12, *by heart* has all its constituents being used in a figurative way. To begin with, *by heart* does not mean the literal heart-it actually refers to committing something to memory. When we are told the clever animals had ‘the entire song by heart within a few minutes, what it implies is that they are able to commit the song to memory. The two constituents *by* and *heart* both have a figurative dimension, which is what makes the multiword unit semantically semi-transparent.

Example 13: *by heart*

-had the entire song **BY HEART** within a few minutes

-the cleverer ones began to learn the commandments **BY HEART**

-learn the seven commandments **BY HEART**. After much

-set to work to receive the new maxims **BY HEART**. Four legs good two legs good

-when they had first got it **BY HEART**, the sheep developed

*To and fro* as example 7 shows is a semi-transparent multiword unit since its constituents **to** and **fro** also function together in a figurative sense as the original meanings of *to* and *fro* is totally divorced from the current meaning of the multiword unit *to and fro*. The constituents of this multiword item conjure up a picture of intense movement between two points. The sentences in example 7 clearly illustrate this claim.

In the three instances of usage listed below *to and fro* suggests some kind of intensity. For example, sentence one in (example 7) reads: ‘the ducks toiled continuously in the sun’ which means they work very hard. The aspect of hard work in this respect contrasts sharply with the original meaning of the constituents which is movement between two points. Finally, collocates of *to and fro* also help us to deduce the meaning of the multiword unit. The verbs of action such as *toil* and *pace* are the examples of collocates that help in the sense disambiguation of the item.

#### 4.3.4 Opaque Collocations

These are multiword units whose meanings are not the sum of the original meanings of the constituents and have no literal counterparts, examples from the data are: *nevertheless*, *drawing room*, *took to her heels*, *turn black into white*, *fallen on evil days*, *set aside* and *windfalls*. These multiword items are not interpretable without considering contextual /cultural knowledge that belies these multiword units (Carter, 1998). What follows is a discussion based on these items.

In example 2, there are illustrations from the data, which reveal that *nevertheless*; is semantically opaque. The compound word has no relationship whatsoever with the original meanings of the constituents such as (*never/ the / less*). Unless one is conversant with the English culture, it is not possible to tell what the word means. Another point is that the context of usage helps in unraveling the meaning in the

sense that; collocates before and after *nevertheless* indicate that a contrary opinion was being advanced. In sentence three, Napoleon's strategy is contrasted with feelings about Fredrick. Needless to say, the compound word is meaningless outside the environment of its most common collocates.

The next item to be considered is *windfall*. The word means fruits that have fallen from a tree especially the first ones to ripen in any season. In the data the word *windfall* is synonymous with apples as the examples below show. Now, the word is considered to be semantically opaque taking into account that its meaning can only be unraveled in the light of the cultural aspects of the English language. However, in the data collocates of the multiword unit *windfall* help in deducing the meaning of the word. In example 14 sentence one, we are told the orchard is littered with *windfalls*. Here, the meaning can only be explained if a connection is made with the words 'the early apples were ripening' which are collocates of the word *windfall*. In that instance, the word is semantically opaque. In example 13 sentence two, the word is still opaque because without considering the culture or prior encounter with the word *windfall*, it is impossible to deduce the meaning of the word. When the usage of *windfall* in sentence number three is considered, it is found to be transparent as there is talk of *windfall apples* which is contrasted with the main crop of apples so that one can guess these are the apples that ripened first. These sentences are listed below.

Example 13: windfall

-the early apples were now ripening and the grass of the orchard was littered with

**WINDFALLS**

-the order went out that all the **WINDFALLS** were to be collected and brought to the harness-room for the use by the pigs.

-without further argument that the milk and the **WINDFALL** apples (and also the main crop of apples when they ripened)

Another opaque item is the idiom *took to her heels* which means escaped. In the data it is used thrice and once again it is impossible to tell the meaning of the multiword unit outside the English cultural context. In sentence 1 example 14, collocates; *galloped away* suggest what the word means. Similarly, in sentence 3; collocates *defend* and *minute* help the in determining the meaning of the idiom.

Example 14: took to her heels

-and the next moment she *took to her heels* and galloped away and into the field.

-gave up trying to defend themselves and *took to their heels* A minute later, all five of them were in full flight...

Two idioms namely *turn black into white* and *fallen on evil days* were also found to be opaque. They have been analyzed in spite of the fact that they have only been used once in the data. This is because unless cultural considerations are made, the

meanings of the two units cannot be established. A prior encounter with the two multiword units is necessary in order to deduce the meanings of these idioms. *Turn black into white* means to deceive or trick as in example 15 below where Squealer can lie convincingly as in ‘he *could turn black into white*’ while *fallen on evil days* means to get into misfortunes or to be broke as in ‘he *had fallen on evil days*.’ Example 16 below sheds more light on this claim.

Example 15: turn black into white

-The others said of Squealer that he could **TURN BLACK INTO WHITE**. These three had elaborated the teachings of Old Major.

Example 16: fallen on evil days

-he had been a capable farmer but of late, he had **FALLEN ON EVIL DAYS**. He had become much disheartened after losing his money...

The other opaque collocation is *set aside* which means to reserve. When it said in the data that the pigs had *set aside* the harness room as their headquarters, the meaning is has to be inferred from collocates of this phrasal verb. Otherwise one may get the impression of the piece of land being lifted up to be placed somewhere which is clearly absurd. The word *set aside* is an opaque phrasal verb so that unless one considers the cultural factor, it is impossible to unravel the meaning. Example 17 below shows how the phrasal verb is used in various sentences.

However the opacity of the word is lower in degree compared to the other members in the opaque category since in the five instances of its usage, its principal collocates are the following nouns: grazing ground, harness-room, corner, paddock and barley. These nouns assist in the sense disambiguation of this phrasal verb.

Example 17: set aside

- previously be intended to be **SET ASIDE** as a grazing ground
- the pigs had **SET ASIDE** the harness-room as their headquarters
- the talk of **SETTING ASIDE** a corner of the pasture for the superannuated animals
- to **SET ASIDE** the small paddock behind the orchard as a
- now that the small field beyond the orchard had been **SET ASIDE** for barley ,it was rumoured that a corner

Finally, the compound word *drawing room* is another opaque multiword unit because it means a sitting room but not a place where drawings are made. The word is fully opaque since there is no connection between the original meanings of the constituents and that of the multiword unit. Example 18 below illustrates this clearly. In the sentences below, *drawing-room* has no relationship with drawings but refers to a sitting room. Unless someone has familiarity with the

English culture, it is impossible to interpret the multiword unit in question as the constituents have a figurative meaning.

Example 18: drawing-room

-he immediately went to sleep on the **DRAWING ROOM** sofa

-the lithograph of Queen Victoria on the **DRAWING ROOM** mantle piece

-took their meals from the kitchen and use the **DRAWING ROOM** as recreation room but also slept on the beds....

-which had been made in the glass cupboard in the **DRAWING ROOM**...

#### 4.4 Relationship Between Restriction and Opacity

An examination of opacity and restriction would not be complete without a discussion on where the two clines meet. From the outset, it must be observed that there are instances where the degree of opacity corresponds to collocation restriction but on a general note, the two aspects of multiword units do affect one another. This is supported by evidence from a restricted multiword unit such as *cowshed* which is found to be semantically transparent. Indeed that observation lends credence to the assumption that no connection exists between opacity and restriction. A multiword unit like *farmhouse* on the other hand is unrestricted on account of its broad collocation range and is also semantically transparent which shows a kind of relationship between opacity and restriction in multiword units.

Any relationship between the two aspects can only be arbitrary so that predictability is totally out of question if these two examples are anything to go by.

The other point is that a multiword unit can be restricted or opaque in different instances of usage. This implies that the same multiword unit can be opaque or transparent depending on the context of usage at any given moment. A good example is *windfall* (see example 13) a compound word. In illustrations number one and two on example 13, the word *windfall* is opaque while in illustration number three it is transparent since the context of usage mirrors the meaning of the multiword unit-*windfall*.

Another example is *windmill* (example 3) where illustrations one and two highlight different aspects. In illustration one, the word *windmill* is used idiomatically and is semantically opaque as in ‘Snowball and his moonshine of *windmills*’. We know that moonshine means foolish ideas in illustration one and therefore has no relationship with *windmill* in illustration two, which refers to the actual power generating machine-*windmill*. In the first instance, the word is opaque in terms of meaning while in the second instance it is a semi-idiom. In the second instance of usage, *windmill* as a semi-idiom corresponds to *windmill* as a semi-restricted collocation.

In conclusion, the analyses found that a language comprises mainly of multiword units. These lexical units behave in the same way as single word lexical items. There are four types of multiword units namely, compound words, phrasal verbs, idioms and prefabs. Prefabs are the most numerous units followed by compound words while phrasal verbs and idioms are third and fourth respectively in terms of frequency of occurrence. Another point is that all multiword units may be gauged using a cline in terms of restriction and opacity. These clines are not mutually exclusive categories but represent the varying degrees of restriction and opacity so that along a continuum of some kind so that a given multiword unit is opaque, transparent, restricted or unrestricted only when compared with others. In short, it is simply a question of what degree of opacity or restriction a given multiword item exhibits.

It should also be noted that the analyses identified the types of multiword units in the data and categorized them into four categories. The items were then sampled and 8 units were the ones subjected to the restriction cline. The sampled units were then examined in terms of their principal collocates and the analyses found that collocates of multiword units are mainly verbs and nouns although other word classes such as adjectives are also represented. Also worth noting is the fact that the opacity cline involved examining the multiword units in terms of semantic opacity. The conclusion drawn is that the more opaque a multiword unit is the harder it is to deduce its meaning. Conversely, the more transparent a multiword

unit is the easier it is to disambiguate it. These observations and others are discussed more intensively in the next chapter.

## CHAPTER FIVE

### SUMMARY OF FINDINGS AND RECOMMENDATIONS

#### 5.0 Introduction

In this section, a summary of the findings of the study is provided followed by recommendations. Suggestions for further research are given and finally a conclusion is given.

#### 5.1 Summary of Findings

This study sought to examine the collocation patterns and degrees of opacity in multiword units with illustrations from George Orwell's novel *Animal Farm*. The study has three objectives.

##### 5.1.1 Types of Multiword Units

The first objective identifies and describes the types of multiword units in *Animal Farm*. Four types of multiword units are extracted from the novel and they are: Compound words, phrasal verbs, idioms and prefabs. The analyses identify a total of 447 multiword units from the data in the following proportions: 121 compound words, 98 phrasal verbs, 63 idioms and 165 prefabs. The multiword units are entered into a table and their percentages calculated. Thereafter, a pie chart is generated from those percentages. The table and pie chart show that prefabs are more numerous than any other multiword unit followed by compound words, phrasal verbs and idioms in that order.

The study also reveals that although multiword units in the data are categorized into four groups, there is much overlap between the said categories so that it is difficult to place some of the units in a particular group. Another point is that the study had proposed to consider five categories of multiword units (Schmitt, 1997). However, only four types of multiword units are identified in the analyses. Compound words fall into three sub-categories i.e. hyphenated, open and fused compounds. Two sub-types of phrasal verbs are also identified and they are the separable and the inseparable phrasal verbs. The idioms in the analyses are classified into either institutionalized or uninstitutionalized sub-categories so that prefabs are the only group of multiword units that are not sub-divided further since they are all highly routinized in language.

### **5.1.2 The Collocation Restriction in the Multiword Units**

The second objective is concerned with exploring and describing the patterns of collocation in the multiword units identified in stage one. This is achieved by subjecting the multiword units to a collocation restriction cline, which has four continuums namely, unrestricted collocations, semi-restricted collocations, familiar collocations and restricted collocations. At this point ten multiword units in the four categories of multiword units are purposively sampled so as to be subjected to the restriction cline. It is, however, observed that most of the sampled multiword units have a low frequency of occurrence in the data. Consequently,

only the units with a frequency of six or more times are subjected to the restriction cline and these were 8 in number. There are three unrestricted multiword units, one semi-restricted multiword unit, three familiar collocations and one restricted collocation. The restrictedness of the multiword items is established by determining the principal collocates and then considering the affinity of a given multiword unit for particular collocates. The multiword unit is the nodal item around which collocates cluster. In addition, a maximum of five words to the right and left of the node are considered. If the multiword unit has a preference for certain collocates, it ranks higher than the previous category in the restriction cline and vice versa. Moreover, it is found that most collocates of the multiword units are verbs and nouns while very few collocates are adjectives or adverbs. It should be observed that; only content words like verbs and nouns are qualified to be termed collocates as opposed to functional words like prepositions since this is a lexical study.

### **5.1.3 Opacity in the Multiword Units**

The third objective considers the degree of opacity in the multiword units. This involves the subjection of 18 multiword units chosen from the purposively sampled multiword units randomly since it is impossible to subject all 40 of them to the opacity cline. The opacity cline has four continuums, which are the transparent collocations, semi-idioms, semi-transparent collocations and the opaque collocations. The degree of opacity is established by considering whether

the multiword units are straightforward in meaning or not. Those in which one can 'see through' are transparent; while the opaque ones are the ones in which one cannot not 'see through'. A given multiword unit is ranked higher or lower in the opacity cline depending on its semantic interpretability.

The study observes that opacity is a scale with various points along which various lexical items can be ranged. It is not an absolute entity but on the contrary, it is just a scale for conveniently describing the multiword unit in terms of semantic complexion. The research identifies four transparent collocations, the key consideration being their literal usage. There are four semi idioms and these items are the ones used in a figurative sense. Also, the study has two semi-transparent collocations and seven opaque collocations. The study further determines that a given multiword unit can be judged to be opaque or transparent depending on the way it collocates in a particular instance. Therefore, the range of collocates is deemed a mandatory consideration if the right meaning is to be assigned to multiword units in a given body of language. Another point is that the research observes the opacity of multiword units does not depend on the category. In fact, the likelihood of a compound word being opaque is similar to that of idioms, phrasal verbs or prefabs. Hence, opacity is an aspect that depends greatly on the context of usage as well as the type and range of constituents for a given multiword unit but not its designation as a compound word or idiom as the case may be. The underpinning culture with regard to a given multiword unit is also

observed to be necessary in the interpretation of the opaque and semi-transparent multiword units. Something else worth noting is that the opacity and restrictedness of multiword units does not have a one to one correspondence except in a few cases but even here; it is the exception rather than the rule.

The findings of the research are hereby highlighted. To begin with it is found that multiword units are very common in a given corpus or body of language and the types are four in number. However, some studies notably Schmitt (1997) describe five types i.e. compound words, phrasal verbs, idioms, fixed phrases and prefabs. These four types according to the research are: compound words, phrasal verbs, idioms and prefabs. Prefabs are the most numerous categories of multiword units while the least numerous categories are those of idioms. Another point is that there are three types of compound words, which are distinguished by their structure. These are: hyphenated, open and fused compound words. With regard to phrasal verbs there are two classes, which are the separable and inseparable phrasal verbs. Idioms are divided into two types on the basis of whether they are institutionalized or not. All prefabs however, are institutionalized; hence, they cannot be subdivided. The other observation is that multiword units collocate with other words and behave just like single lexical units. Multiword units are also viewed as examples of complex collocations whose most frequent collocates are verbs and nouns.

In any given corpus multiword units act as nodal items around which collocates cluster. The principal collocate is the one with the greatest attachment to the multiword unit. The multiword units collocate with content words such as verbs so that functional words like prepositions do not matter in a lexical study like the present research. The character of these collocations is such that the multiword units form patterns, which are described in terms of restrictedness. The restrictedness varies from one multiword unit to the other and is described using a cline Carter (1998). The cline runs from the unrestricted to the restricted collocations with semi-restricted and familiar collocations as intermediate categories.

Something else worth noting is the fact that multiword units form an opacity cline, which refers their increasing semantic complexity when subjected to a semantic cline. The semantic cline runs from transparent collocations to the opaque collocations with semi-idioms and semi-transparent collocations in between. The opacity of multiword units depends on the degree of familiarity of the users of a given language with the underpinning culture. This is because a given multiword unit may be transparent in one instance but opaque in another depending on how it collocates in the two contexts. However, there is no one to one correspondence between opacity and restriction in multiword units. In fact the opacity or restriction of multiword units do not depend on their categories but on the contrary, they depend on their contexts of usage as well as on their constituents.

## 5.2 Recommendations

This study focuses on the multiword units in *Animal Farm* as a representation of the multiword items in other contexts. The other contexts include plays, newspapers, magazines and songs among others. The study reveals specific things about multiword units in language such as the link between culture and the interpretability of multiword units. Another revelation is the role of context and the collocation range in the assigning of meaning to multiword units. However, this research is library based and merely makes recommendations to researchers and language practitioners on how best to handle multiword units in language teaching and learning contexts. The study aims at inspiring more research in multiword units especially in the indigenous languages of Kenya. Therefore the recommendations below target prospective researchers in multiword units, language teachers, authors, publishers and syllabus designers.

First literary works such as novels are rich in multiword units so that these resources ought to be employed more and more in language teaching and learning. This is because multiword units are the core of any language and should be the focus of language teaching to ensure learners are proficient in a given language. Another recommendation is that, on account of the varying degrees of opacity in multiword units, language practitioners must devise ways of ensuring their charges interact sufficiently with multiword units to ensure they are able to recognize and use them appropriately in real life contexts. This especially concerns

opaque multiword units. Thirdly there is need to explore the unique patterns in multiword units in order for language users to manipulate language appropriately. Another point is that syllabus designers, teachers and publishers must veer off their traditional preoccupation with grammar and accord lexis the central place it deserves in pedagogy if their efforts are to bear fruit it being common knowledge that multiword units make up a third of language. In addition, teachers must expose learners to a lot of material in the target language so as to make it possible for learners to interpret the multiword units encountered later in the learning process with ease. This is especially important in second and foreign language teaching. The strong connection between the interpretability of multiword units and cultural background makes it necessary to expose learners to real time language since according to Sinclair's Idiom principle and the Lexical approach of Michael Lewis; words are produced not as single lexical units but as strands of preconstructed phrases. The real time language will enable them to encounter the multiword units in ideal contexts of usage.

### **5.3 Areas for Further Research**

This research focuses primarily on the multiword units in a novel written in the native English lexicon i.e. *Animal Farm*. However, multiword units are phenomena of all languages in all contexts of usage, which implies that further research may be carried out in the following areas: To start with, an analysis of the multiword units in a second language speakers' lexicon can be investigated.

Secondly, the presentation, teaching and learning of multiword units at primary school or secondary school level is another area that may be studied. Thirdly, the multiword units in the bible and other Christian literature may also be studied to reveal their frequency of occurrence and the uniqueness of the patterns they form compared to those in non-religious literature. Also, the multiword units in Kenyan newspapers and magazines can be examined to see the extent to which these collocations are used as well as the way they have been used. To this end, a comparison may be made with newspapers from countries in which English has a first language status. Another area in which prospective researchers can venture is the exploration of the multiword units in texts written in any of the indigenous languages of Kenya. Finally, the researchers may consider investigating the multiword units in radio and television broadcasts or the multiword units in the academic writing of Kenya University students since these phenomena have been studied in other countries but as yet, such studies have not been carried out in Kenya.

#### **5.4 Conclusion**

This research is quite informative for it reveals unique aspects of multiword units in general. The source text has very many multiword units and this leads to the observation that novelists and writers in general make generous use of multiword units. The research assumes that multiword units are not given the attention they deserve in language teaching and learners are expected to 'wade through the

waters of language' and acquire them incidentally. This explains why the language proficiency of many graduates from our institutions is so pathetic in as much as a proper manipulation of multiword items is lacking. Lexical studies like Howarth (2007), equate language competence with the ability to use multiword units properly. The onus is, therefore, on language practitioners to devise ways of enabling learners to acquire multiword units systematically and consciously rather than incidentally. This will result in reasonable degrees of competence especially in second and foreign language learning contexts.

## BIBLIOGRAPHY

Atkins, J. (1971) *George Orwell*. Great Britain: Calder and Boyar Limited.

Bejoint, H. (2000). *Modern Lexicography: An Introduction*. Oxford: Oxford University Press

Calder, J. (1987) *Animal Farm and Nineteen Eighty- Four*, Philadelphia: Open University.

Carter, R. (1998) *Vocabulary: Applied Linguistics Perspectives*. London and New York: Routledge.

Coady and Huckin (1998) *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press.

Collie and Slatter (1995) *Literature in the Language Classroom: A Resource Book of Ideas and Activities*. Glasgow: Bell and Bairn Limited

Cruse, D.A (1989) *Lexical Semantics*. Great Britain: Cambridge University Press.

Crystal and Davy (1969) *Investigating English Style*, UK: Longman Group

Culler, J. (1975) *Structuralist Poetics: Structuralism Linguistics and the Study of Literature*. London: Routledge.

Gairns and Redman (1990) *Working with Words*. Cambridge: Cambridge University Press.

Gardner, D. (2007) *Validating the Construct of a Word in Applied Corpus – based Vocabulary Research: A Critical Survey*. Oxford: Oxford University Press

- Hoey, M. (1995) *Patterns of Lexis in Text*. Oxford: Oxford University Press
- Hoey, M. (2005) *Lexical Priming: A New Theory of Words and Language*.  
London: Routledge
- Howarth, P. (2007) 'How Conventional is Academic Writing.' Leeds: University of Leeds.
- Howarth, P. (2007) 'A phraseological Approach to Academic Writing.' Leeds: University of Leeds.
- Leech, G. N. (1969) *A Linguistic Guide to English Poetry*. London and New York. Longman Group.
- Lewis, C. S. (1994) *Studies in Words*. Cambridge: Cambridge University Press.
- Lewis, M. (1997) 'Implementing the Lexical Approach: Putting Theory into Practice.' UK: Language Teaching publications
- Malmkjer, K. (1991) *The Linguistics Encyclopaedia*. London: Routledge
- Mburu, M. W. (2003) 'Lexical and Phonological Errors in the Spoken English Language of Pre-school Teachers in Classrooms in Kasarani Division:'  
Kenyatta University, Kenya.
- Meyers, J, (1975) *A Reader's Guide to George Orwell*. London: Thames and Hudson
- Milroy. (1987) *Observing and analyzing natural language: A Critical Account Of Socio Linguistic Method*. Oxford: Basil Blackwell

- Ndung'u, R. W. (2002) 'The use of Synonyms by Second Language Learners of English: Patterns, Lexical Strategies and Pedagogic Implications.' Kenya: Kenyatta University.
- Nyamasyo, E .A. (1992) 'A Corpus Based Study of Grammatical and Lexical Characteristics of the Writing of Kenyan Pre-University Students.' Lancaster: University of Lancaster
- Ongele, E. A. (2005) 'Examination of the Suitability of a Corpus Driven Instruction to the Teaching and Learning of Phrasal Verbs in Kenya. Kenya: Kenyatta University
- Oxley, B.T. (1967) *George Orwell*. London: Evans Brothers Limited
- Palmer, F. R (1988) *Semantics*. Cambridge: Cambridge University Press
- Patton, M.Q(1990) *Qualitative Evaluation and Research Methods*: California: Sage Publications
- Richards, J. et al. (1987) *Longman Dictionary of Applied Linguistic*. UK: Longman Group.
- Schmitt and Conklin (2007) 'Formulaic Sequences: Are they processed more quickly than Non-formulaic Language by Native and Nonnative speakers?' Oxford Journals: Oxford University Press.
- Schmitt, N. and M. McCarthy (1997) *Vocabulary: Description, acquisition and Pedagogy*. Cambridge: Cambridge University Press
- Singleton, D. (2000) *Language and the Lexicon*. London: Arnold.

- Vintar, S. (2007) 'A Lexical Analysis of the Elan Slovene – English parallel Corpus.' Slovenia: University of Ljubljana
- Wykes, D. (1987) *A Preface to Orwell*. New York: Longman Group Limited
- Zwerdling, A. (1974) *Orwell and the Left*. Britain: Yale University Press.

## APPENDICES

## Appendix1: Compound words identified in the data

Compound words			
Bedroom	Newspaper	hawthorn bush	recreation room
beforehand	nightshade-	hayfield	ringleader
birch saplings	berries	headquarters	rolling-pin
birthday	nosebags	hen-houses	razor blade
blacksmithing	nose-rings	high-pitched	seed corn
blood-curdling	offset	hill side	false step
growls	outdoing	hind leg	farm buildings
bodyguard	outhouses	drawing room	farmhouse
brainwork	outside	dressing-table	field-drains
brass medal	outside world	drinking pool	five-barred gate
horse brasses	setback	easy-going	foodstuff
horsehair sofa	sheepfold	gentleman	footprints
human beings	side whiskers	fairness	lin seed
knot-hole	fore leg	moonshine	lithograph
lawsuits	fore lock	overstrain	looking-glasses
lifetime	gate-way	overthrow	lump sugar
mantelpiece	general	overtook	main road
masterpiece	assembly	paintbrush	mangel-slicer

elm tree	governess-cart	plough land	flagstaff
midsummer's eve	graveside	pop-holes	limestone
mankind	harness-room	public-houses	swill-bucket
electric-milking	cartloads	worn-out	tale-bearer
machine	chaff-cutter	outweigh	taproom
everlasting	circular saw	overcame	tiptoed
sidelong look	cooperative-	overgrown	title deeds
sly-looking	enterprise	override	overslept
snowdrifts	countryside	underfed	brass-studded collars
spring sitting	cowshed	vegetable	breakfast
store-shed	crowbars	garden	carpetbag
sucking-pig	cupboard	watchword	cart track
woodland	dogcart	well-being	dressing-table
	drawback	windfalls	drinking pool
	drawing room	windmill	

**Appendix2: Phrasal verbs identified in the data**

Phrasal verbs				
adapted to	bound to	broke out	carried off	cleared up
attended to	break up	broken of		come after
blown off	breeding	built up	carried on	come into
come to	up	fall down	carried	let out
come upon	made off	fence off	through	lodge in
comes on	painted out	flung down	given out	lurched across
conjured	crept back	fluttered up	go off	pass on
up	crept out	gaining on	go up	passed out
make up	cutting off	galloped	go upon	peered in
mingle	die down	away	keep up	pick up
with	drive out	gazed at		plucked up
moved of	drag down	give up	keep off	stir up
paid up	run away	written down	kicked off	stove in
privy to	run up	made for		strolling across
put forward	rushed out	set aside	lay down	take in
put on	sat down	set against	trace out	take flight
put to vote	saving up	spell it out		taken away
puzzle out	sent out	sprang off	trodden on	taken out
ran	worked out	spread		fled by

through		across	turned out	worn off
read out		stand up		
rubbed off		slipped out	went off	
		sold off	went on	
			went out	
			wore on	
			pass it off	
			settled down	

**Appendix3: The idioms identified in the data**

Idioms		
bitter winter	root cause	keep close to
a spirit of license and indiscipline	scream your lives out	keep the breath in our bodies
an air almost of amusement	set the animals minds	lead the way
biding his time	at rest	lead you astray
lost heart	set the table in a	lost hope
by the hour	shared such sentiments	marshal his thoughts
changed their tune	friendly feelings that	no better fate
cleared his throat	subsisted	short and to the point
close on his heels	grudged no effort	stroke of good fortune
command a view completely	held aloof	work to the last atom of our strength
at ease	in the same strain	take advantage of
rob us of our repose	in the teeth of every difficulty	turn black into white
take life	iron discipline	were at hand
they opened fire	to drink a toast	while the going was good
they stopped short	to laugh to scorn	with a knowing air
they take refuge	to lose an hour's sleep	word had gone round
tired out but proud	took a place	
fallen on evil days	took to their heels	

struck dumb with surprise Contain their rage courage returned to them crystal clear	cruel knife deadly silence deep silence hung over the farm house did not square with this draw attention	driving hard bargains emptied to the dregs nothing could be further from the truth put the whole matter in its perspective
---	--	---

**Appendix4: The prefabs identified in the data**

Prefabs		
on the contrary	in fact	At last
on the following	in front	at least
once again	in order to	contrary to
simply as	in spite of	for example
so far as	in the fact that	for instance
sooner or later	in the meantime	from now onwards
such and such	last of all	henceforward
thereafter	moreover	if anything
throughout	nevertheless	in addition
until now	now that	all the same
within	nowadays	and so on
according to	as though	and yet
after that	as usual	apart from
at first	as well	as for
of course	above all	as soon
barrel of	in a moment	afterwards
baying of	tears rolling down his	depth of character
better and better	cheeks	efforts were rewarded
bunches of	if need be	filled nearly to the brim

by heart	in league with	with sand
cast a very ugly look	in reality	flights of
catch a glimpse of	wisps of	flocks of
cry of despair	with one accord	for an instant
cryptic remarks	without cause	for nothing
now that	wasted time	for the benefit of
suspicious still lingered	way round every	for your sake
system of thought	never mind	for hours on end
slept as they had not slept	no way	force of gravity
before	nodded in complete	frightened though they
starve to death	agreement	were
now that	on the spot	in ruins
at any late	out of control	it was no use
for days at a time	out of reach	it was natural
not even	difficulty	like a radish
in between	well enough	little more to say
a brood of	by the time	long ago
a bushel of	by and large	none of the
a dozen	seat of honour	not even
a game of	seen no more	his cheeks
a long way away	set forth	out of the way
a lump of	to and fro	over the edge

a pack of	traces of	pleasure of your old age
a piece of	twice as thick	plenty of
a pile of	utmost care	point of honour
a pint of	as quietly as	rain or shine
a shout of triumph	as surely as	shivers down all the
a tour of inspection	out of spite	animals spines
absolutely necessary	lying in ambush	side by side
at the top of his voice	just in time	side to side
get rid of	majestic looking	stand it no longer
good-for-nothing	make themselves	as to the effect
head of the table	comfortable	of late
here is my toast	many years ago	by and large
his own advantage	my sight is failing	by the time
hoisting of the flag	reserved for	in agreement
hot and cold	retinue of dogs	stack of
long been expected	round and round	
long live	sack of	
line of attack	some degree	
	something else	

**Appendix 5: The multiword units sampled from each of the four classes of multiword units**

<b>Compound words</b>	<b>Phrasal verbs</b>	<b>Idioms</b>	<b>Prefabs</b>
cowshed	breed up	cruel knife	as soon
bedroom	turn out	iron discipline	in a moment
windmill	given out	bitter winter	for instance
five-barred	settled	deadly silence	side to side
gate	down	on all fours	by heart
farmhouse	put	marshal his	round and
hayfield	forward	thoughts	round
drawing room	set aside	close on his	a lump of
moonshine	drag down	heels	to and fro
nevertheless	fall down	fallen on evil	hot and cold
windfalls	pass on	days	above all
	lodge in	turn black into	
		white	
		word had	
		gone round	

**Appendix 6: The multiword units that were subjected to the restriction/opacity clines**

Compound words	Phrasal verbs	Idioms	Prefabs
Cowshed	pass on	turn black into	as soon as
farmhouse	set aside	white	side to side
windmill		fallen on evil days	round and round
five-barred gate		close on his heels	to and fro
drawing room		cruel knife	by heart
windfalls			
nevertheless			