



University of HUDDERSFIELD

University of Huddersfield Repository

Alsuweed, Muhammad

Accessing Dickens's style as an EFL learner: a corpus stylistic approach to lexical style

Original Citation

Alsuweed, Muhammad (2015) Accessing Dickens's style as an EFL learner: a corpus stylistic approach to lexical style. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/31480/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

Accessing Dickens's Style as an EFL Learner: A Corpus Stylistic Approach to Lexical Style

Muhammad Alsuweed

A thesis submitted to the University of Huddersfield
in partial fulfilment of the requirements for
the degree of Doctor of Philosophy

The University of Huddersfield

November 2015

To my dear parents, my dear wife, my family and my little sons, including the new arrival.

Abstract

This study is based on a corpus of the Charles Dickens Complete Works (the DCC), which was constructed to fulfil the aims of this research. The DCC was compiled to represent *The Works of Charles Dickens* in the *National Edition* (a set of 40 volumes, including the life of Dickens in the last two volumes, which consists of 6,202,886 tokens in total). **This compilation, as the DCC, represents the first complete corpus of Dickens's works.** Employing the corpus stylistic approach was as an underpinning concept, and formed the methodology that has guided the research. **The lens of focus is placed on Dickens's lexicon, in respect to both the lexemes and their relative frequency, alongside the choices of lexis to be found in the context. The rationale for this thesis and value of its aims is primarily the facilitation of non-native English learners' access to these works, through provision of an enhanced aesthetic appreciation of Dickens's style with regards to his semantics and lexical choice. Additionally, the methodology aims to enable the acquisition of vocabulary, while providing learners with training in the reading of complex texts.** The software tools used in the analysis are the *WordSmith Tools 6.0* suite, *AntConc 3.4.4w*, *AntWordProfiler 1.4.0w* and the *Range* programme. The investigation of the DCC was conducted to facilitate Dickens's works to non-native readers by focusing on the lexicon of his works. The analysis reports, amongst others, the *DCC keyword list*; the *DCC Headword List* (with 27,296 headwords); and the *DCC Word Family List* (approximately 102,753), which contains the family members of each headword in the DCC. These lists represent a valuable resource that can serve to facilitate the teaching of Dickens objectively, and through an evidence-based approach. In essence, the lexical knowledge gained from the DCC is intended to advance the reading and comprehension of Dickens's works by non-native readers, and then to contribute towards the development of such learners' level of English language proficiency. **Therefore, this study builds bridges between corpus stylistics and**

second language pedagogy. In the analysis of Dickens's lexical selection, I demonstrate how learners can be assisted to reach the appreciation of Dickens's style in terms of his lexicon and the semantic level of his works.

Acknowledgements

My wish is to convey my deepest thanks to all those who have tirelessly supported and unconditionally encouraged me through this challenging process of writing this doctoral thesis. In particular, my gratitude is extended to my supervisor, Professor Dan McIntyre, for his continuous support, encouragement, critical feedback and observations, and his belief in my work and ability; these have been fundamental to my completion of this thesis. I would further like to extend my gratitude to Professor Lesley Jeffries, for her constant encouragement and compassionate support.

Although at times my progress was challenged, during these periods, the positive and compassionate encouragement from my wife, friends and supervisors has propelled me forward, and helped to stimulate the necessary motivation to continue. Completion of this thesis would not have been possible without their personal and academic support.

Furthermore, I would like to sincerely offer thanks to my wife in particular, for her caring support, and complete understanding during every phase of this thesis's journey. As for my sons, I can only apologise to them for my preoccupation on those occasions when I have been consumed by my work. I pledge that my achievement will be a reward for them also.

I also offer my heartfelt thanks to my dear friends, who have been there to advise and support me along the way: Sheikh Abdul-Karim for his compassion, wisdom and moral support, and Dr Muhammad Rababaa and Dr Shabbir Dastgir for their pertinent advice and encouragement.

The staff members at the University of Huddersfield have been incredibly helpful, offering speedy and intelligent solutions to those issues that have cropped

up during this PhD journey. In particular, I would like to thank the Head of the Department of History, English, Languages and Media, Professor Paul Ward, and the Dean of the School of Music, Humanities and Media at the University of Huddersfield, Professor Martin Hewitt. Finally, my gratitude is offered to the staff members of the International Office, Mrs Lianne Holmes and Miss Kirsten Barker, for their ready support whenever requested.

Copyright statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the “Copyright”) and s/he has given The University of Huddersfield the right to use such copyright for any administrative, promotional, educational and/or teaching purposes.
- ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.
- iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the “Intellectual Property Rights”) and any reproductions of copyright works, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

Table of Contents

Chapter 1 Style, Reading and Dickens’s Works	1
1.1. Introduction	1
1.1.1 Style and stylistics	1
1.1.2 Stylistics and lexical choice	4
1.1.3 Literary text and learning language	6
1.2 Context of the study	11
1.2.1 Stylistics in the literature and language classroom.....	12
1.2.2 Lexical choice as a pillar of language teaching.....	14
1.3 Aims and research questions.....	20
1.4 Definition of commonly used terms in the thesis	22
1.5 The structure of the thesis.....	27
1.6 Conclusion	29
Chapter 2 The Nature of Authentic and Simplified Texts	31
2.1 Introduction	31
2.2 Simplified texts.....	34
2.3 Authentic texts.....	37
2.4 Comparative analysis of authentic and simplified texts.....	39
2.4.1 The lexical profile statistics of simplified and authentic samples.....	44
2.4.2 Word-family lists.....	48
2.5 Lexicon or structure: prioritising the focus	56
2.5.1 Language competence	58
2.6 Conclusion	61
Chapter 3 Stylistics, Corpus Linguistics and Corpus Stylistics.....	63
3.1 Introduction	63
3.1.1 Stylistics and linguistics	63
3.1.2 Possible challenges for non-native readers	67

3.1.3 The historical dimension in Dickens's works	70
3.1.4 The cultural dimension in Dickens's works.....	71
3.1.5 Utilising corpus stylistics to facilitate stylistic analysis.....	72
3.2 Stylistics, its aim and objectivity	74
3.3 Corpus linguistics	81
3.4 Corpus stylistics	85
3.4.1 The theoretical concepts and principles of corpus stylistics.....	90
3.4.2 Opening doors with the corpus stylistic approach.....	93
3.4.3 The difference between corpus linguistics and corpus stylistics.....	102
3.4.4 Using the corpus stylistic approach in teaching literature	104
3.4.5 Limitations of the corpus stylistic approach.....	107
3.5 Conclusion	109
Chapter 4 The Charles Dickens Complete Corpus	111
4.1 Introduction	111
4.2 Collecting Dickens's works	117
4.3 The retitled works of Dickens	120
4.4 Identifying the complete works of Dickens	122
4.5 Dickens's seemingly uncollected works.....	126
4.6 The construction of the DCC.....	128
4.6.1 Selecting texts' sources.....	129
4.6.2 Storing and checking	132
4.7 Description of the DCC.....	133
4.8 The value of the DCC	134
4.9 Conclusion	136
Chapter 5 The Applications of Corpus Stylistics.....	139
5.1 Introduction	139
5.2 Wordlists	141
5.3 Keywords	144
5.4 Collocations.....	149

5.4.1 Defining collocation.....	149
5.4.2 The significance of collocation	155
5.4.3 Exploring collocation through corpus methodology.....	157
5.4.4. The application of collocation in language teaching	160
5.5 Lexical bundles, formulaic expressions and key clusters: interrelated phenomena.....	162
5.5.1 Lexical bundles.....	162
5.5.2 Formulaic sequences.....	167
5.6 Key clusters.....	169
5.7 Semantic prosody.....	170
5.8 Conclusion	174
Chapter 6 Analysing the Charles Dickens Complete Corpus	176
6.1 Introduction	176
6.2 The wordlist frequency of the DCC	178
6.3 Keywords	188
6.4 Collocations.....	192
6.5 Lexical bundles.....	194
6.6 Key clusters.....	196
6.7 Application of grading the task of reading.....	198
6.8 Vocabulary sophistication	201
6.9 Guiding the reading of Dickens’s works.....	205
6.10 Dickens’s lexical selections	210
6.10.1 Analysis of word choice: <i>accoucheur</i>	211
6.10.2 Analysis of word choice: <i>round-aboutedly</i>	212
6.11 Exploring the DCC’s key concepts	213
6.11.1 The <i>Anatomy & Physiology</i> category (B1).....	219
6.11.2 The <i>Unmatched</i> category Z99	223
6.11.3 <i>Light</i> : key concepts in the DCC	227
6.11.4 <i>Darkness</i> : key concepts in the DCC.....	235
6.12 Conclusion	240

Chapter 7 Conclusion	243
Bibliography	249
Appendices	269
Appendix 1.1: Function Words by Paul Nation	269
Appendix 4.1 Texts Manually Inserted into the DCC	271
Appendix 4.2: The Comprehensive List of Charles Dickens’s Works.....	285
Appendix 6.1: The DCC Keywords List	326
Appendix 6.2: Three-Word Lexical Bundles of the DCC	353
Appendix 6.3: Key Clusters in the DCC Compared with the BNC.....	357
Appendix 6.4: The Dickens’s Word List (Sample)	363
Appendix 6.5: The Overused Key Concepts in the DCC with Log-Likelihood above 6.63 (99% accuracy).....	366

List of Tables

Table 2.1 The simplified versions of two of Dickens’s novels	40
Table 2.2 Comparison between the simplified and authentic texts of two of Dickens’s novels	40
Table 2.3 Headwords’ percentage increase between each stage/level at the simplified levels, and at the authentic text level	41
Table 2.4 <i>Oliver Twist</i> – lexical profile statistics for simplified paragraph.....	44
Table 2.5 <i>Oliver Twist</i> – lexical profile statistics for authentic paragraph.....	45
Table 2.6 <i>Great Expectations</i> – lexical profile statistics for simplified paragraph.....	45
Table 2.7 <i>Great Expectations</i> – lexical profile statistics for authentic paragraph.....	46
Table 2.8 The contents of Paul Nation’s BNC/COCA word-family lists.....	49
Table 2.9 <i>Great Expectations</i> – simplified without proper nouns	54

Table 4.1 The Dickens Corpus of the University of Giessen: 43 items	112
Table 4.2 The DCorp's contents: 23 items	113
Table 4.3 Contents of the University of Giessen's Dickens Corpus extracted from other works	115
Table 4.4 Items extracted from <i>Complete Ghost Stories</i>	121
Table 4.5 Contents of the DCC precisely matching the National Edition volumes..	130
Table 5.1 The concordances of the word <i>becoming</i>	144
Table 5.2 The collocates for the word 'write' in the DCC (using AntConc 3.4.4w).	155
Table 6.1 The DCC sub-corpora and word counts	178
Table 6.2 The twenty most frequent words in the DCC	181
Table 6.3 The twenty most frequent lexical words in the DCC.....	183
Table 6.4 The twenty most frequent lexical lemmas in the DCC	185
Table 6.5 The concordance plot for 'I' in the DCC.....	186
Table 6.6 Example from the DCC Word-Family List	192
Table 6.7 Top twenty collocates for 'eye' to the left of the node collocate.....	193
Table 6.8 Key clusters of the DCC compared with the BNC.....	197
Table 6.9 The first two sentences from <i>Oliver Twist</i> : authentic and simplified versions.....	200
Table 6.10 Examples of unconventional spellings found in <i>Oliver Twist</i>	202
Table 6.11 Paul Nation's baseword lists.....	202
Table 6.12 The DCC's lexical profile statistics.....	204
Table 6.13 Lexical profile statistics for <i>Oliver Twist</i>	205
Table 6.14 Headwords in each part of <i>Oliver Twist</i>	207
Table 6.15 Lists resulting from the analysis of <i>Oliver Twist</i>	208
Table 6.16 Ranking Dickens's fifteen novels chronologically.....	208
Table 6.17 Ranking Dickens's fifteen novels according to their headword diversity	209
Table 6.18 The DCC's first 20 key semantic concepts	215

Table 6.19 The 21 discourse fields at the semantic hierarchy in general levels (Wmatrix)	217
Table 6.20 The 100 most frequent items in the <i>Anatomy & Physiology</i> category (B1) of the DCC	219
Table 6.21 Antonymous key-concepts in the DCC	227
Table 6.22 Key semantic concepts of the <i>Light</i> (W2) category in the DCC.....	228
Table 6.23 The ten most frequently occurring collocates for the word 'light'	230
Table 6.24 How 'light' collocates with 'eye' in the DCC	231
Table 6.25 Key semantic concepts of the <i>Darkness</i> (W2) category in the DCC	235
Table 6.26 The 10 most frequent collocates with 'dark' in the DCC	236
Table 6.27 'Dark' and its 99 collocates in the DCC	236

List of Figures

Figure 2.1 A view of the authentic passage of <i>Oliver Twist</i> in the AntWordProfiler 1.4.0w's File Viewer and Editor Tool	51
--	----

Table of Abbreviations

AWL:	Academic Word List
BDPF	Brewer's Dictionary of Phrase and Fable
BE06:	British English 2006
BNC:	British National Corpus
CALD:	Cambridge Advanced Learner's Dictionary
CID:	Collins Italian Dictionary
COCA:	Corpus of Contemporary American English
COED:	Concise Oxford English Dictionary
CCALD:	Collins COBUILD Advanced Learner's Dictionary
DCC:	Charles Dickens Complete Corpus
EFL:	English as a foreign language
EPUB:	electronic publication
F-LOB:	Freiburg–LOB Corpus of British English
GSL:	General Service List
HTML:	hypertext markup language
LDCE:	Longman Dictionary of Contemporary English
LL:	log-likelihood
LOB:	Lancaster-Oslo/Bergen Corpus
OALD:	Oxford Advanced Learner's Dictionary
PDF:	portable document format
POS:	parts of speech
Roget's II:	Roget's II: The New Thesaurus
SLA:	second language acquisition
TXT:	plain text

Chapter 1 Style, Reading and Dickens's Works

1.1. Introduction

This study presents an investigation of how learners of English can be assisted in their reading of Charles Dickens's works as authentic texts. Besides the acquisition of vocabulary, engaging with original unsimplified texts requires an understanding of stylistic techniques. Therefore, the two aims that this study seeks to achieve are: firstly, to investigate Dickens's lexicon with the intention of assisting learners to acquire and develop the vocabulary necessary in order to read and comprehend Dickens's works; and secondly, to raise the awareness of **learners** so that they can appreciate Dickens's style by focusing on the semantic aspects of his work through consideration of his word choice and collocations. I use corpus stylistic methods to facilitate the achievement of the aims of this study by focusing on Dickens's work, as it is widely appreciated for its literary value, alongside the notion that reading complex authentic texts is considered to be a benchmark for skilled readers with a high level of language proficiency. **In this section, the nature of style and stylistics, stylistics and lexical choice, and the use of literary texts for language learning will be introduced.**

1.1.1 Style and stylistics

Style manifests as a result of linguistic selection. *Style* in language is a comprehensive term encompassing the linguistic organisation and patterns at a range of levels – semantic, syntactic, phonological and pragmatic – which in unison comprise the language in its structure and complexity. In general terms, as Fischer-Starcke (2010) suggests, style in language can be considered to consist of grammatical rules (the syntactic level) and lexical items (related to the semantic and pragmatic levels). Fischer-Starcke (2010), for instance, views *style* as being the 'lexical and grammatical patterns in a text that contribute to its meaning' (Fischer-Starcke 2010: 2). The meaning of texts is investigated through these two levels, that is, the lexis and grammar, as Jeffries (2006)

asserts that '[t]he study of **semantics** in linguistics has historically been targeted at the two components of language: the **lexis** (vocabulary) and the grammar' (Jeffries 2006: 156; emphasis in original). The primary concern of stylistics is the consideration of linguistic style in texts (Jeffries & McIntyre 2010). Berg and Martin-Berg (2002) consider that 'style is the choice made by a speaker or writer among the various equivalent expressions available in a language for communicating a given potential content' (Berg & Martin-Berg 2002: 175). Moreover, they 'define the adjective "stylistic" in the broadest possible sense as the application of the above definition of style to the act of reading and interpreting the resultant choice of expressions' (Berg & Martin-Berg 2002: 175). The focus of stylistics is as a systematic approach that explores the language of the text, while simultaneously excluding from its approach any aspects that fall beyond the text. Therefore, the stylistic reading of texts is grounded in the notion of the text as a set of selections employed by the author through a range of language levels inclusive of the semantic, syntactic and phonetic domains. The result of these stylistic selections is that the text functions in a specific manner to produce meaning. Stylistics is an approach that considers the language of the text as a means to achieving an end, that is, to comprehend and appreciate the style embedded within the language, which can be described as a group of selections at the level of lexicon and the structure of the language, and how they interrelate within a given context. Stylistics necessarily returns to the linguistic aspects and properties of the text. The objective of locating several of these properties should concentrate on the constituent units of the text, and the manner in which they emerge and relate. Wales (2011) views stylistics as a discipline centrally concerned with the formal, and particularly linguistic features of texts, together with their functions for readers in elaborating meaning and interpretation. Wales (2001) also notes the increasing emphasis in stylistics on social and cultural contexts, through which it can be determined what texts are selected, which features are noticed and how they might be interpreted.

Discussing the definition of stylistics, Thornborrow and Wareing (1998) confirm that 'stylistics is concerned with the idea of "style", with the analysis of literary texts, and with the use of linguistics. "Style" is usually understood within this area of study as *the selection of certain linguistic forms or features over other possible ones*' (Thornborrow & Wareing 1998: 2; emphasis in original). They proceed to confirm that '[a] stylistic analysis of the styles of these writers could include their words, phrases, sentence order, and even the organization of their plots' (Thornborrow & Wareing 1998: 2). Stylistic features, as defined by Bussmann (1996), are 'the repetition or mixing of **elements of style** and, therefore, on the particulars of the grammatical form ... on the vocabulary ... or on the structure of the text' (Bussmann 1996: 1135; emphasis in original). Bussmann (1996) further considers the *element of style* to include '[a]ny linguistic element that determines the **stylistic features** of a text', seeing that 'any linguistic phenomenon can have a stylistic function'. These stylistic elements/features include 'phonetic elements of style, (**alliteration, phonostylistics**), lexical elements of style (**nominalization, archaism**), morphological elements of style (genitive ending 's), syntactic elements of style (sentence complexity, length of sentence), and textual and pragmatic elements of style (types of **cohesion, theme-rheme, thematic development**)' (Bussmann 1996: 355; emphasis in original).

Naciscione (2010) **emphasises** that attaining skills and competency in stylistics is a time-intensive process, and one that must be maintained moving forward rather than merely being learned, while Leech and Short (2007) highlight that learning about style necessitates the identification of pertinent features of language, with this selection criteria being dependent upon 'the purpose we have in mind' (Leech & Short 2007: 12).

If stylistics is concerned with the analysis of a range of linguistic features in respect to their functionality in the text, this study will focus on one main aspect of stylistic devices: the lexis in the text, that is, Dickens's lexicon. One of the measurements to be revealed by investigating Dickens's lexicon is the lexical diversity in his works, as 'lexical diversity variables have been applied to many areas of linguistic investigation.

These include journalism; lexical innovation and loss; literary style; language register ...' (Malvern et al. 2004: 6).

The rationale for this study's focus on style is because it is an approach to the interpretation of text (e.g. see Picken 2007: 18). Wales (2011) underscores this focus on the linguistic features and interpretation of a text, since

[t]he goal of most stylistic studies is to show how a text "works": but not simply to describe the FORMAL features of the texts for their own sake, but in order to show their FUNCTIONAL significance for the INTERPRETATION of the text; or in order to relate literary effect or themes to linguistic "triggers" where these are felt to be relevant.

(Wales 2011: 400, **small capitals in original**)

Jeffries and McIntyre (2010) reformulate this perspective by stating that 'analysing style means looking systematically at the formal features of a text and determining their functional significance for the interpretation of the text in question' (Jeffries & McIntyre 2010: 1).

1.1.2 Stylistics and lexical choice

This study examines the lexicon of Dickens's works by means of corpus stylistics, thus highlighting the semantic values contained within the lexicon. The lens will focus on Dickens's lexicon, as semantic units through which the meanings of the text are manifested; for example, the focus will include the most frequently occurring words, and how their emergence statistically leads to their significance stylistically, considering how different forms convey differing messages.

The reasons why this study focuses on lexical style are as follows:

- 1 – The choice of lexical items is a stylistic device, which thus accommodates learning about the style used in a given text by a specific author;
- 2 – In second language acquisition (SLA), the role of vocabulary is crucial in reading and comprehension; and

3 – It also assists learners in realising ‘lexical competence’ (see Section 2.5.1).

Focusing on the lexical items of Dickens’s works as the source from which the meanings are created will simultaneously lead to the consideration of the context in which these lexical items arise, as they do not solely or intrinsically embody all meanings independently, but rather such lexical items can create other meanings from the contexts in which they occur. Widdowson (1975) discusses what he refers to as the ‘significance’ and ‘value’ of lexical items, as well as the concept of ‘texts’ and ‘discourse’, clarifying the relevance of context and language in use. Nevertheless, it is beyond the scope of this study to attempt to provide a detailed analysis for the entire works of Dickens at all linguistic levels, that is, the semantic, syntactic, phonological, and pragmatic patterns. As Short (1996) cautions, ‘[a]nalyzing a long *novel* in close stylistic detail could take a *lifetime*’ (Short 1996: 255; emphasis added). Rather, the focus will be restricted to the lexicon of Dickens, which will be investigated from two angles: (i) as a main component in language acquisition (considering *grammar* as the other main component) which learners need in order to comprehend a given text; and (ii) as a stylistic device, since word choice can be employed to manifest a specific effect on the readers’ interpretation.

Corpus stylistics offers a valuable approach for the analysis of literary texts, ‘allowing for the quantification of recurring linguistic features to substantiate qualitative insights and vice versa’ (O’Keeffe 2006: 50).

Sinclair (2004: 148) advocates the use of ‘computational resource’ for the ‘identification of lexical items’ as the

impact of corpus evidence on linguistic description is now moving beyond the simple supply of a quantity of attested instances of language in use. It is showing that there is a large area of language patterning ... that has not been properly incorporated into descriptions; this is the syntagmatic dimension, of co-ordinated lexicogrammatical choices.

(Sinclair 2004: 148)

Language has a range of features that can be depicted consistently, such as its morphology and phonological characteristics, together with the pragmatics, semantics and syntax (Gass & Selinker 2008). These linguistic dimensions are at the same time the levels of stylistic analysis. This study concerns itself with the semantic/lexical analysis of Dickens's text, as will be detailed below.

Semantic analysis is an important element in stylistic analysis (Mahlberg & McIntyre 2011: 206), due to its significant impact on the process of constructing meaning, which is not limited to an indicial factor; nevertheless, the lexical items which **comprise** the text play a vital role in the process of meaning creation. In researching lexical items, the focus can be directed to the following:

1. **Words and their differing forms**, particularly where their sense is significantly affected by such change.
2. **Collocations**, as there are certain words that can rarely be utilised unless they are accompanied by certain other words. These collocations can be observed with regards to a particular topic, or for a specific author.

1.1.3 Literary text and learning language

The main aim of this thesis is concerned with the stylistic analysis of particular literary works, namely those of Charles Dickens. The rationale behind the focus on style is, as Berg and Martin-Berg (2002) suggest, that the concept of style can span the gap that is found between the study of foreign languages and literature which can erode the potential of the classroom, the coherence of the curriculum and wellbeing of the teachers. If we are to promote literature as 'the highest form of expression' (**Gilroy & Parkinson 1996: 213**), then this is due to the manner in which it reveals the mechanics of the language, **which is one of the aims** of teaching foreign languages. Moreover, **Behnam (1996) asserts that** rather than expecting non-native learners 'to elicit some intuitive responses for which little evidence can be drawn', they may focus more practically on the more measurable device of style in order to assist them in 'activating

a general analytic skill which can serve as a first step towards [the] ultimate interpretative responses expected from readers' (Behnam 1996: 12). Berg and Martin-Berg (2002) suggest that the concept of style allows learners to identify, explore, understand and ultimately appreciate the linguistic mechanics and strategies with which the writer engages in order to create form and meaning. **Through understanding the function of the style of the language, learners are supported to comprehend literary text as a complex discourse.** Brumfit and Carter (1986) consider that it

is almost the only "context" where different varieties of language can be mixed and still admitted. Any deviation from norms of lexis and syntax in legal documents would be inadmissible ... any non-literary linguistic form can be pressed into literary service. Writers will exclude no language from a literary function.

(Brumfit & Carter 1986: 8–9, cited in Simpson 1997: 18)

Investigating lexical items offers insight into the field of stylistic studies while shedding light on the distinctive features of a particular text or author. This semantic level of analysis followed by stylisticians is through conducting a linguistic study of literary texts, applying their preferred linguistic analysis, and then employing statistics to reveal that stylistics offers intratextual standards to literary critics that can be relied upon in order to arrive at an objective interpretation.

The value of the stylistic approach lies in the fact that it linguistically informs the use of literary texts as learning materials in the classroom. Although style within language can manifest in any type of genre, literary texts are a common example where stylisticians and researchers aim to explore these different styles. For learners of English as a foreign language (EFL), English literature is regarded as one of the authentic sources through which they can acquire, consolidate and extend their English language proficiency. In order to explore style in language, the language itself is appraised when deployed in actual use. Naciscione (2010) asserts that literary texts of the modern era have gained recognition as a significant medium for language acquisition and

awareness raising, which is applicable not merely in the context of 'lexical and grammatical accuracy', but additionally to enhance 'literary awareness and stylistic sensitivity' (Naciscione 2010: 205). The ability to gain insight through stylistics enables the ability to extend understanding that then penetrates 'the surface meaning of words', to illuminate that which is embedded 'between the lines of spoken and written discourse' (Naciscione 2010: 205). As style is consistently manifest in language and can be better identified with a context, the study of style will result in particular attention being paid to the facet of context, considering the links bridging context and language, and perceiving literature's contextual components as a facilitator to interpretation (Naciscione 2010). Naciscione (2010) observes that the lens of discourse stylistics is focused on how language is used and those stylistic characteristics that are deemed vital for the comprehension of and engagement with extended chunks of texts, offering the examples of **Carter and Simpson (1989a)**, Carter (1996), Cook (1994), Emmott (1997), McCarthy and Carter (1994), and McRae (1987) (Naciscione 2010: 206). Naciscione (2010) then cites McRae and Boardman (1984) in stating that the capability of comprehending and appreciating literary texts reflects a key component of learned native speakers' communicative competency, and one that can be underscored as the ultimate aim of non-native learners and their tutors (Naciscione 2010: 30). Stockwell and Whiteley (2014) see that stylistics embodies an approach that limits its focus to investigating the linguistic characteristics of the text, together with the manner in which meanings are constructed. This perspective differs from the manner in which literary critics treat (literary) texts by way of, for instance, examining the aesthetic impact, or other effects of the texts. Stylisticians **argue** that literary critics can become more systematic and objective when accommodating stylistic approaches while dealing with (literary) texts.

This study is intended to meet non-native readers' needs by furnishing them with more targeted advice on lexical and stylistic usage, aimed towards non-native speaking readerships. Reading and the appreciation of a text's style enable learners to

acquire knowledge, while supporting their self-confidence and facilitating language development. Style is not fundamentally in contrast to lexis and grammar; however, I believe that learning style can scaffold learners in understanding and appreciating literary texts. This study's intention is to help learners understand Dickens's style by focusing on one of his style aspects, namely his lexical choice.

The inclusion of the stylistic approach in assisting learners to read and interpret Dickens's works is an attempt to provide learners with a precise, practical, systematic and objective approach for the successful achievement of such a goal. It also extends beyond the literature criticism approach in teaching literary texts, which is heavily reliant on the intuition of learners. Kramersch and Kramersch (2000) emphasise the manner in which the field of stylistics has informed the teaching of literary texts in language classes, leading to elevated students' awareness regarding the style employed in language use, while simultaneously paying attention to the social context in which the language operates. They offer exemplifying studies to support their assertions such as Carter and McRae (1996), Carter and Simpson (1989a), Cook (1994), Fowler (1996), Kramersch (1993), Short (1996), Simpson (1993), Toolan (1998) and Widdowson (1975 1992) (Kramersch & Kramersch 2000: 569). The corpus stylistic approach will be employed in this thesis to achieve both aims, firstly by identifying and grading the lexicon required to perform a specific reading task related to Dickens's work, and secondly through identifying patterns related to Dickens's choice of lexicon as one of the stylistic devices.

There are two observations concerning the use of the stylistic approach. Firstly, this study does not only concern itself with the primary/dictionary meanings of the lexical items in Dickens's works, as reliance on dictionaries alone cannot always account for meaning. Hence, the shifting of the lexical items' meanings from the basic/primary to the contextual should facilitate the learners' ability to capture the message intended by the author. These dictionary definitions are also inadmissible because the entries of the reference books to hand are not always suitable for studying the change or

evolution of words and their meanings. Such a task (i.e. establishing the accurate meanings of Dickens's work, considering his time of writing and usage) should be based on the context and usage, which is consequently what the stylistician practises.

Secondly, the stylistic approach facilitates in the discovery of the internal topical structure of the text. Simpson (2000) indicates towards this concept, confirming that 'examining the internal topical structure ... [is] reflected by the repetition of key words and phrases, [which] provides insights into the organizational patterns favored by' Dickens (Simpson 2000: 293). It is of value to highlight that both of these observations can be achieved through a text-based approach, which is the one employed in this study.

Hernández (2011) addresses the potentiality of literary texts in language teaching from a functional direction and claims that such texts appear to represent an opportune resource to facilitate the development of 'literary comprehension and sensibility', while elevating communicative competency as literary texts (i) presents language in authentic use, (ii) offers numerous opportunities 'for the expression of ideas, opinions, and beliefs', and (iii) functions as a starting point for productive writing. Moreover, literature embodies the potential to extend the 'psycholinguistic aspect of language learning' through its emphasis on 'form and discourse processing skills', while promoting the extension of vocabulary and supporting reading competencies, with text appearing to offer an ideal medium through which to explore 'the stylistic features of an author and the characteristics of a period' (Hernández 2011: 235–6).

Hall (2005) believes that the specific language of texts to be engaged with should be stimulating to both the learners and their language teachers. This is particularly evident within the context of literary texts as authentic texts and the level of difficulty, where the challenges can be related to a number of areas such as the 'plot, cultural references and text world', as well as in the linguistic sense, where non-native learners may encounter difficulties in respect to challenging 'vocabulary, style or register, syntax, rhetoric, genre or discourse organisation' (Hall 2005: 130). Likewise, the domain

of modern stylistics (on occasion referred to as discourse stylistics), typically aspires to extend beyond merely a perfunctory breakdown of a text's linguistic characteristics towards a 'functional interpretation', whereby the interpretations of linking characteristics are highlighted, and the readers discover these through their focus on texts where such features are embedded (Hall 2005: 130).

In defending the value of using the stylistic approach in teaching literature, Widdowson (2008) states: '[h]erein lies its educational value – for it offers an alternative to the traditional teaching of literature. Rather than being the passive recipients of the second hand interpretations of literary critics, students can be enabled (empowered even) to take the initiative and engage actively and directly with literary texts themselves' (Widdowson 2008: 302). Semino (2011) asserts that the pedagogical uses for stylistic analysis have been of interest to stylisticians in respect to native tongue, second language and foreign language contexts (see Widdowson 1975; Widdowson 1992; Watson & Zyngier 2007). The application of classroom-based stylistics is considered to offer the opportunity of embracing both language and literature tuition, while representing an alternative to the prevailing literary critical approach that is typically employed in international literature classrooms. Kettemann and Marko (2004) argue through detailed descriptions how the use of corpus-based stylistics in the classroom can raise awareness in 'different dimensions', 'namely language awareness, discourse awareness, and methodological and metatheoretical awareness' (Kettemann & Marko 2004: 170).

1.2 Context of the study

This section explores the application of stylistics in the literature and language classroom, before making a case for this study's focus on lexical choice as a fundamental component of language teaching.

1.2.1 Stylistics in the literature and language classroom

To begin with stylistics, Zaro Vera (1991) describes the traditional method of teaching literature, which shares many similarities with the context of the English Language and Translation Department at Qassim University, Saudi Arabia (my own institution), where this study aims to contribute towards the development of literature tuition. The teaching of literature is concerned with the 'study of plot, characterization, motivation, etc, and also, sometimes overwhelmingly so, a study of the author and their historical and literary background' (Zaro Vera 1991: 164). Such teaching depends on 'lecturing' and the students' ability to record class notes, engage in feedback and ultimately sit examinations on the lecture material and readings from the curriculum, and while the reading medium is extensive, together with excerpts explored within the classroom, the tuition is typically teacher-centred, with the entire process tending 'to be transmissive and product-centred in respect to the outcome being predictable' (Zaro Vera 1991: 164).

Adopting a stylistic approach can inform the teaching of literature from a linguistic perspective by offering significant precision and objective procedures (e.g. Leech 2008; Widdowson 1975), with Zaro Vera (1991) suggesting that 'understanding about language has direct consequences for understanding the meanings of the work of literature, that interpretations are often not simple intuitions but language functions, which differentiates modern stylistics from previous critical approaches like the «New Criticism» or «Practical Criticism»' (Zaro Vera 1991: 167). In his response to the question of why stylistics is used in second and foreign language contexts, Hall (2007) asserts that the 'use of literary texts is often advocated as a means to enhance proficiency in reading, vocabulary growth and cultural knowledge, if not indeed, in more traditional systems' (Hall 2007: 4). However, in its defence, McEnery and Wilson (2001) highlight that the 'whole concept of style rests on the notion that authors choose to express their ideas using certain linguistic resources in preference to others, which logically must be measurable to some degree' (McEnery & Wilson 2001, cited in Murphy 2007: 67).

Naciscione (2010) believes that applied stylistics **enables the broadening of** comprehension **regarding** how stylistics are employed, generating greater understanding of the nature and function of stylistic features, while adapting the perspectives and rules that form the foundation of language teaching and curriculum design. The focus on the 'applied value of a stylistic approach' is essential when creating teaching resources to support the study of texts of a literary and non-literary nature (Naciscione 2010: 207). **In developing a discourse stylistic model, the features that can be maintained, as suggested by Carter and Simpson (1989b) include firstly that the territory for the stylistic analysis of texts can be located beyond the exchange of a single conversation or at the sentence level, while secondly those wider contextual properties of texts that impact on how they are described and interpreted are considered (Carter & Simpson 1989b: 12).**

Thornborrow and Wareing (1998) indicate towards the existing domain of discourse stylistics, which aims to progress the strands of 'intra-textual analysis and text/reader interaction' (Thornborrow & Wareing 1998: 169). Furthermore, they assert that discourse stylistics allows researchers to consider texts as they are found in the 'real world', as opposed to them being viewed as merely 'independent aesthetic artefacts', thus recognising the dynamic relationship that exists between texts and their readers by claiming that meaning emerges from a range of features inclusive of how the text has been formally assembled and the context in which it is being read (Thornborrow & Wareing 1998: 168). The contemporary notion of discourse stylistics is of a field where literary texts are studied through the lens of discourse analysis, which is concerned with interpreting the 'semantic and stylistic relationships in text' (Naciscione 2010: 16). Naciscione (2010) considers, for instance, that '[a]nalysis of cohesion (Halliday & Hasan 1976: Chap. 8) is central to discourse stylistics because it reveals semantic and stylistic links' (Naciscione 2010: 16).

Naciscione (2010) **posits** that discourse analysis 'encourages interpretation in discourse through [the] exploration of meaning to integrate the study of language and

literature' (Naciscione 2010: 15), providing the examples of Carter (1997), Carter and McRae (1996), **Carter and Simpson (1989a)**, Cook (1992), Cook (1994), Emmott (1997), Leech and Short (2007), Short (1988), and Widdowson (1992). Moreover, Naciscione (2010) believes that '[d]iscourse stylistics deals with [the] interpretation of semantic and stylistic relationships in text' and that the '[a]nalysis of cohesion (Halliday and Hasan 1976: Chap. 8) [is] central to discourse stylistics because it reveals semantic and stylistic links' (Naciscione 2010: 16). Therefore, the principles and practices of teaching and learning vocabulary need to be reconsidered within the context of a discourse-based approach to language use (Naciscione 2010: 208).

The rationale behind focusing on Dickens's lexicon is to promote lexical competence, as an aspect of communicative competence. According to Hudson and Eigsti (2003), '[l]exical competence is a speaker's facility with the content lexical items in a language, and includes overall vocabulary size as well as ease of lexical access' (Hudson & Eigsti 2003: 1). Lexical competence is, as Velasco (2007) suggests, 'part of communicative competence' (Velasco 2007: 166), with *communicative competence* itself referring to 'a speaker's grammatical competence plus knowledge of the rules and conventions governing the accurate, appropriate and effective use of the language in a wide range of social settings' (Brown & Miller 2013: 88). According to Meara (1996), 'lexical competence is at the heart of communicative competence' (Meara 1996: 35).

1.2.2 Lexical choice as a pillar of language teaching

After considering the use of stylistics in the teaching of language and literature, the lens is now turned towards the role of lexis in language learning, and the development of a rationale for this study's focus on lexical choice as a fundamental facet of language learning. Vocabulary represents an **essential** component in the learning of a foreign language due to its semantic value, which represents the core of communicative competence amongst the users of that language. However, knowing semantic elements in an abstract manner (e.g. wordlists) does not facilitate successful communication with

other users of that language, as such meaning that can be learnt from wordlists only reflects the narrow or limited interpretation as it appears, for instance, in dictionaries. Lexical items can have multiple definitions, unlike a single word in a list which will inevitably be restricted to fewer meanings. Hence, the transition from teaching and learning vocabulary lists to teaching and learning vocabulary as it functions in discourse is intended to assist in identifying the implied sense within that given context. This indicates that teaching vocabulary via wordlists has shifted towards contextual teaching that supports the identification of the contextual meanings of said words. Such definitions are determined by the associated discourse (**co-text**), with the sentence focusing on the investment of the semantic relationship between lexical items in order to create a given meaning to a specific lexical item. The complex structure of sentences, words as a group or a larger unit or chunk of the language, contribute to specifying the intended meanings, which is therefore a facet of the discussion on the teaching and learning of vocabulary from the discourse analysis perspective.

There is a difference between teaching vocabulary in isolation (i.e. vocabulary lists), and in the context of sentences or discourse. Teaching vocabulary in isolation in the form of lists is characterised as a useful approach to assist in focusing on the meanings of the individual words (e.g. see Nation 2001). It is then teaching for meaning without addressing the context or how such words might function within a text. Conversely, teaching vocabulary through the context of sentences or discourse is intended to determine accurately the meanings of these lexical items by way of exclusion of other possible meanings found within the dictionary that the context does not support. This approach of teaching lexical items in their discourse also pays greater attention to the context in which a lexical item occurs, how it functions in the sentence in terms of the grammar or discourse, and what collocations it is associated with.

The discourse analysis perception of vocabulary is significantly different from traditional approaches to teaching vocabulary, that is, learning through wordlists. This does not imply that there is negligible value in teaching vocabulary through lists or

sentences since it is a requirement at the very early stages in language learning, even if those words are isolated from the context in which they typically occur. After that stage, where the teaching of vocabulary adopts the list and sentences' method, vocabulary can be taught in longer contexts (i.e. via discourse) which can be considered as a discourse analysis approach or corpus stylistics approach to teaching vocabulary; it **is** a stylistic approach in the sense that it focuses on how meanings are created in the text by selecting specific lexical items to convey a particular message.

Some of the differences that distinguish the teaching of vocabulary via wordlists and as a discourse analysis approach include the following points regarding the latter:

1. It is based on authentic texts that reflect the actual occurrence of language and communication in a natural manner. Therefore, these types of texts are not adapted, graded or modified for specific educational purposes.
2. It highlights the status of the vocabulary and how it functions to generate meanings, that is, the meaning linked with different contexts. The nature of its relation with other vocabulary and how they contribute to the production of the meaning of the text, the production of content and the significance of coherent texts are all sought in the discourse analysis (see Carter (1998: 220) for the issue of vocabulary in discourse). As Carter (1998) observes, there is 'a distinct shift from examining lexical items at the level of the orthographic 'word' or in the patterns which occur in fixed expressions towards a consideration of lexis in larger units of language organization' (Carter 1998: 79).

Contemporary linguistics concurs on the point, which confirms that words have no meaning in isolation, unless they function **in a context to create meaning** (Halliday 1999). This indicates that a word within a discourse has relationships with other words, which combine to determine the sense of the word. Hence, **it could be argued that** the word **may** lose its potency when located outside of the context.

Although the *elements of style* can manifest in all the dimensions of the language (e.g. semantics, syntax and phonology) they can for the sake of analysis be approached from one of the following directions. Stylisticians typically investigate style by addressing a specific aspect, for instance the lexical selection, sentence structure or deviation. The **selection of lexical items** is employed as a discourse producer that has a vast reservoir of lexical items from which (s)he can select some certain ones to express a specific meaning. At this level, the research is directed towards highlighting the reasons that justify such choices, and the possible effects that they create. The principle of choice or selection is considered to be one of the characteristics that stylistic research is interested in. Language contains numerous items which can be utilised to compose a virtually unlimited number of phrases and sentences. Therefore, the issue to address in some stylistic analyses is to explore in-depth the reasons and signs related to such selection, of say one clause or sentence over the other, or the reasons behind preferring particular lexical items, to illuminate the implied effect behind such choices. The selection process thus remains a helpful measurement by which we may detect the uniqueness of a text over another or a writer over another, through studying the style of lexical variety (e.g. see Jeffries 2010).

Asserting the significance of the lexicon **from an SLA perspective**, Gass and Selinker (2008) state that in recent research into SLA, less focus has been applied to lexicon in contrast to the remaining components of language, although shifts in this landscape are unfolding with rapidity. Nevertheless, there are a range of justifications to support the notion that lexis has real relevance in SLA, with there being potential for lexicon to represent the most essential component of language for learners, **while** prominent corpora of errors suggest that non-native learners' lexical errors are the most prevalent (Gass & Selinker 2008).

Shuy (2001) suggests that 'writers have rather high levels of consciousness and control over vocabulary choices but considerably less consciousness and control over their grammar, spelling, or punctuation patterns. Discourse style is another language

feature of which most speakers and writers have little or no conscious awareness or control' (Shuy 2001: 450). The focus on the conscious choices of vocabulary carried out by Dickens intends to illuminate how Dickens deliberately conveys his meaning. In this study, the investigation of Dickens's lexicon is a variant of discourse analysis in the sense that it deals with authentic texts as they occur naturally without amendment, modification or simplification, to serve educational purposes, for instance. Moreover, authentic texts offer a high degree of credibility.

The significance of lexicon is also recognised from the perspective of the lexical approach, a language teaching approach. Lewis (1993) states:

The Lexical Approach develops many of the fundamental principles advanced by proponents of Communicative Approaches. The most important difference is the increased understanding of the nature of lexis in naturally occurring language, and its potential contribution to language pedagogy.

(Lewis 1993: vi)

Listing, the '[p]rinciples and implications of the Lexical Approach', he adds that:

- Language consists of grammaticalised lexis, not lexicalised grammar.
- Although structural patterns are acknowledged as useful, lexical and metaphorical patterning are accorded appropriate status.
- It is the co-textual rather than situational elements of context which are of primary importance for language teaching.
- Grammar as structure is subordinate to lexis.

(Lewis 1993: vi–vii)

In the identical argument for the case of lexical items, Lewis (2008) reports that John Sinclair stated unequivocally at IATEFL 1996 that a lexical mistake often causes misunderstanding, while a grammar mistake rarely does so (Lewis 2008: 16). Wilkins (1972) asserts that 'without grammar very little can be conveyed, without vocabulary nothing can be conveyed' (Wilkins 1972: 111). Gass (1988b) supports the view that learners find lexical errors to be more troublesome than their grammatical counterparts, 'noting that grammatical errors generally result in structures that are understood, whereas lexical errors may interfere with communication' (Gass 1988b, cited in Gass & Selinker 2008: 449).

Lewis (1993) employs the nomenclature *context* to refer to features that relate to the situation and *co-text* to refer to the linguistic surroundings. Essentially, context can be considered as the existing world knowledge that interlocutors hold (Brown & Miller 2013), while co-text refers to the associated linguistic surroundings of a section of discourse (Burke 2010) or the broader ‘textual context of some piece of text’ (Brown & Miller 2013: 114). Bex, Burke and Stockwell (2000) consider that co-text refers to that which surrounds the analysed text while context can encompass ‘the social and cultural backgrounds, which bring a text into being, and the social, cultural and cognitive positionings of those readers who interpret the text and give it meaning’ (Bex, Burke & Stockwell 2000: i). Therefore, co-text can be said to be the linguistic and textual environment of a lexical item, such as the concordance lines displaying part of the co-text of a lexical item, while context is the non-verbal environment in which a word can be found. The significance of investigating lexical items in their co-texts lies in, what Green (2000) confirms as being ‘[m]eaning partly derived from surrounding sentences’ (Green 2000: 63, emphasis in original), since ‘words do not stand in a one-to-one relation with concepts, and do not possess core or invariant meaning ... language acts like an indexical writ large: it indicates rather than encodes. Linguistic items are only prompts for certain cognitive activities’ (Green 2000: 51). The studying of Dickens’s lexicon will be conducted on its co-text in this thesis with the help of corpus linguistic techniques, which assist in the examination of words in co-text (see Gilquin 2010: 169).

Having underscored the importance of lexical choice as a key theme in language learning, the next section introduces the study’s aims and the research questions that have been formulated in order to respond to these aims. My intention is to establish the importance of focusing on lexicon in my study for three important reasons: firstly, the choice of lexicon is a stylistic device that is harnessed to analyse a text stylistically; secondly, from the SLA perspective it is an essential component in language learning; and thirdly, from the approach of language teaching (i.e. *the lexical approach*) language consists of *grammaticalised* lexis, *not* lexicalised grammar (Lewis 1993: vi–vii), and thus

the focus should be directed towards the lexicon for stylistic analysis, acquiring English as a second/foreign language and teaching English a foreign language.

1.3 Aims and research questions

This study seeks to achieve two aims: firstly, to investigate Dickens's lexicon, with the intention of assisting non-native readers of Dickens's works in reading and comprehending the targeted work by advancing their lexical competency; and secondly, to raise non-native readers' stylistic awareness so that they can appreciate Dickens's style, triggered by his lexical choices. For the first aim, it is important to quantify and investigate the employed lexical items in order to gain knowledge of the most important aspect of the style of literary texts, which is lexicon, besides estimating the richness of Dickens's lexicon if it is to be compared with other writers of reference corpora (e.g. the richness and diversity of vocabulary and how verbs, nouns and sentences are harnessed to discover other aspects of the style characteristics of the writer). As for the second aim, one of the prominent stylistic features of a text is the lexicon employed in the discourse. In literary texts, for instance, lexicon is exploited in the formulation and production of specific discourse. The implication of such exploitation is in the delivery of a specific message in literary discourse/works. Therefore, through the examination and studying of the richness of the vocabulary of literary texts, this leads to the identification of one of the most distinctive features of the style, since the lexical items utilised in creating the texts function as the building blocks used by the writer to establish and construct a sentence which reflects their character and uniqueness amongst others. Regarding the stylistic awareness, Naciscione (2010) states:

The importance of stylistic awareness lies in development of the learner's perception of language in use and his or her response to it. *Stylistic literacy* is a functional ability to use stylistic skills competently for applied purposes and activities. It is a skill that will help to apply language more purposefully and effectively.

(Naciscione 2010: 207)

Addressing ‘the importance of dealing explicitly with questions of style’, Argamon, Burns and Dubnov (2010) consider that ‘[i]n text analysis, style is key to understanding the feelings and social relationships expressed in a text ...’ (Argamon, Burns & Dubnov 2010: vi–viii). Naciscione (2010) states that to nurture an awareness of style is vital to comprehend and appreciate the use of language in texts of both literary and non-literary natures, and that through placing emphasis on how language is used and the features of style, empirical research has sought to combine the approaches of the linguist and the literary critic in a particular interpretation and consider both the poetic function and **form** (Jakobson 1960, cited in Naciscione 2010: 15). Moreover, post-Jakobson a linguistic perspective to literature has resulted in a great many approaches and conducted research, with one of the most significant advances being the momentum of considering language from the discourse position (Naciscione 2010: 15).

Systematically assisting learners in the acquisition of the required lexicon (as will be shown in Chapter 6) and raising their stylistic awareness will eventually contribute to advancing their communicative competence in using English. It is my view that communicative competence is the ultimate goal of learning a second or foreign language. VanPatten and Benati (2010) indicate that communicative competence is inclusive of *grammatical competence*, which refers to the ‘knowledge of the linguistic forms and structure of language’ (VanPatten & Benati 2010: 72), or in other words the knowledge of the lexicon and grammar. It should be noted that scholars differ in what sub-competences they place within the communicative competence domain besides *grammatical competence*, but they generally include *discourse*, *pragmatic*, *sociolinguistic* and *strategic* competences (VanPatten & Benati 2010).

Therefore, the questions that I will answer in this study to achieve the stated aims are:

1. What are the Dickensian lexemes that learners need to know in order to provide 100% coverage of Dickens’s entire works (the DCC)?

2. How can learners be equipped to appreciate Dickens's style with reference to his lexical choices?

1.4 Definition of commonly used terms in the thesis

The key terms utilised in this study include *word*, *lexicon*, *lexical item*, *lexeme* and *collocation*. To place these terms in order from the specific to the more general, *word* can be defined first. Matthews (2007) defines a *word* as 'the smallest of the units that make up a sentence, and marked as such in writing' (Matthews 2007: 436–7). Singleton (2000) indicates that the 'linguists' attempts to provide a general characterization of the word have made reference to quite a wide variety of possible defining properties' (Singleton 2000: 1) and discusses '[t]he main lines of these different approaches' (Singleton 2000: 6), which include the orthographic approach, phonetic approach, phonological approach, semantic approach and grammatical approach (see Singleton 2000: 7–10). To illustrate the terms as they will be employed in this study, considering the related areas here of SLA, corpus linguistics and semantics, they are defined as follows:

Lexicon, as Singleton (2000) indicates, 'basically means "dictionary", and it is the term used by linguists to refer to those aspects of a language which relate to words, otherwise known as its *lexical* aspects' (Singleton 2000: 1). In defining *lexicon* Baker, Hardie and McEnery (2006) refer to the term as a 'list of words', suggesting that

[i]n **corpus linguistics**, this usually refers to a list of words held on computer, sometimes with extra information about each word on the list. But one may also speak of "the lexicon of a language". This refers to all the words that exist in that language – which cannot ever be fully listed in practice, since new words are being created all the time.

(Baker, Hardie & McEnery 2006: 108; emphasis in original)

Lexicon in corpus linguistic terms can also refer to the number of tokens in a corpus (see Baker, Hardie & McEnery 2006: 162). According to Lewis (2008: 8), lexicon does not merely consist of single-word items, but also features multi-word items of seeing so that 'the other categories ... provide the novelty and pedagogic challenge' of

words (single lexical items), collocations, fixed expressions and semi-fixed expressions (Lewis 2008: 8). Therefore, when referring to Dickens's lexicon, this signposts the entirety of lexical items employed in Dickens's works in contexts that could be any one of the four categories mentioned above.

Lexical item, also referred to as *full, lexical* or *content words*, points towards those words that embody significant meaning, even where they are found beyond the boundaries of context (Singleton 2000). Lexical items are commonly contrasted against function words (also referred to as *grammatical, empty* or *form words*), as the latter typically have little or no meaning when employed independently and perform a broadly grammatical purpose (Singleton 2000). Examples include, *the, to, for* (see Appendix 1.1 for a list of 320 function words suggested by Nation 2001).

Lexeme, also called *word form*, is 'a lexeme's concrete representatives or realizations' (Singleton 2000: 5). This concept will be helpful when discussing Dickens's word families. *Word family* refers to 'related forms or as an abstract unit which is realized by one or other of these forms as the linguistic environment demands - calls to mind the concept of the phoneme and its allophones' (Singleton 2000: 5). Jeffries (2006) considers that the nomenclature of *lexeme* is typically employed to indicate 'a collection of forms that are grouped together under the same denotation' (Jeffries 2006: 158), with Singleton (2000) concurring that the concept of the word as either a class of forms or 'an abstract unit' is embraced in the term *lexeme*. An associated terminology is the *citation form of the word*, which is employed in association with the meaning of a text and its semantic content, with the distinction between content and function words being often drawn. With respect to lemmas, corpus software can lemmatise lexemes by categorising various words or their forms (Matthews 2007). Thus, the two terms 'headword' or 'lexeme' can be considered as being interchangeable when applied to vocabulary or keyword analysis: they are both in their basic forms that can extend to other possible forms. Further potential terms that can represent the same phenomena include the *canonical* and *dictionary forms* and *citation form*. Kennedy (1998), referring to *lemmas* and

headwords, states that ‘it is normal in corpus studies to list under the same headword or lemma the inflectional variants’ (Kennedy 1998, cited in Knowles & Don 2004: 70).

Collocation, according to Baker, Hardie and McEnery (2006), refers to the phenomenon that particular words have a higher likelihood of occurring in association with other certain words in particular contexts. Therefore, a collocate is a word which occurs with frequency ‘within the neighbourhood of another word’ (Baker, Hardie & McEnery 2006: 35–6). Stylistically, Halliday (2007) asserts that collocation ‘is a highly effective stylistic device’ (Halliday 2007: 153), stating:

The collocation of words is the basic formal relation in lexis. It is extremely important for the study of the language of poetry, since poets, and writers in general, draw their effects in part from the interaction of familiar with new collocations; and the creation of new collocations, interacting with other linguistic features, is a highly effective stylistic device.

(Halliday 2007: 153)

On the other hand, and from an SLA perspective, Schmitt (2000) considers that ‘[c]ollocation is an advanced type of vocabulary knowledge’ (Schmitt 2000: 89; emphasis in original), and thus demands specific attention in order to be learned effectively. Assisted reading tasks are suggested in this study besides learning vocabulary, as their discourse promotes the acquisition of collocations.

Carter (1998) suggests that a ‘main argument throughout is that lexical items in discourse require to be constantly interpreted and re-interpreted by the language user and that, when analysts move beyond constructed examples to a consideration of real texts, the “values” of lexis become of considerable significance’ (Carter 1998: 80).

Lexical competence reflects a significant portion of language proficiency, as it indicates that the learner can comprehend the meanings of lexical items and their interrelations, and their receptive or productive use, in addition to other morphological and grammatical facets of the lexical items. One of the manifestations of the relationship between lexical items is the collocational formations.

In his objection to Chomsky's **concept** of *linguistic competence* as a framework **that shines light on the linguistic knowledge native speakers possess of a language, which he distinguishes from linguistic performance (i.e. the manner in which a language system is employed in enable communication)**, Hymes (1972) proposes the notion of *communicative competence*. In noting that the linguistic and grammatical rules known by speakers are not sufficient to continue to establish the right level of communication in the community, Hymes (1972) introduces the need for communicative competence, which assists in establishing successful communication that is appropriate for the context. Thus, this term has centred on the social component of language and the function of social practice in the completion of the role of theoretical linguistic competence (see Bagarić & Djigunović (2007) for a detailed discussion of Chomsky and Hymes's positions on competences).

The communicative competence impact extends beyond the linguistic theoretical discussions to affect other fields in Applied Linguistics and the teaching of EFL. The communicative approach as a teaching methodology is informed by the concept of communicative competence, leading to situational/contextual teaching and to language learning for specific purposes. The curriculum design is also affected by these important concepts (see Stelma (2010) for a detailed account of the impact of communicative competence on language teaching). Therefore, the study of Dickens's lexicon falls within this category, as investigating the lexical items from a functional perspective and considering their role in communication and the use of language are related to the notion of communicative competence (both functionally and communicatively).

Lewis (2008) discusses the nature of fluency and its dependency on the breadth of the learner's lexicon, which is itself reliant on the volume and quality of the linguistic input that the learner has been exposed to. The potentiality of investigating lexical items from the discourse analysis perspective stretches beyond the conventional boundaries of teaching vocabulary (e.g. wordlists, and meanings at the level of sentences).

McCarthy and Carter (1994) indicate towards Halliday and Hasan's (1976) 'description of lexical cohesion' in investigating the vocabulary patterns as 'research into vocabulary in extended texts' (Halliday & Hasan 1976, cited in McCarthy & Carter 1994: 65). The concept of 'lexical cohesion' in the context of studying vocabulary, as in discourse analysis, in a way refers to the 'repetition of words', and such repetition can produce 'certain basic semantic relations between words in creating *textuality*, that property of text which distinguishes it from a random sequence of unconnected sentences' (McCarthy & Carter 1994: 65; emphasis in original). This offers a different perspective on vocabulary than other disciplines such as the semantic or syntactic approaches which focus on the isolated meanings that relate to a specific context or the grammatical or structural characteristics, that is, the function of the word in a sentence, without broadening the study of that meaning into an extended text or discourse. So, it could be argued that the status of vocabulary in the discourse exceeds the isolated significance or meaning, or its simple grammatical function, to arrive at the function of production and receiving the discourse, which contribute to its features and the manner in which it can be discussed. The status of the vocabulary in the discourse represents the associated semantic, structural and communicative elements. As discourse analysis focuses on the concept of the text and elements leading to it, it is natural to consider the vocabulary from a new viewpoint, albeit one that depends on the semantic and lexical data; but the purpose in discourse analysis is the investment of these semantic relations in producing a coherent discourse, and thus does not restrict its search to the semantic relations without their associated discourse context. The application of stylistic analysis techniques is intended to support learners at advance levels of proficiency, due to the fact that the purpose of such methodology is to assist **them** to comprehend the functionality of how words create discourse, and how they can contribute towards identifying the characteristics of the discourse and its type. Therefore, the purpose is not only to teach and acquire the vocabulary, but rather to understand how they function and create meanings in a given discourse.

Behnam (1996: 18) reports that Carter (1986) encourages a transition from the solely text-based tradition towards a discipline that is more oriented to the social and communicative domains, thus moving away from the analysis of 'literature as text' towards that of 'literature as discourse' (Carter 1986, cited in Behnam 1996: 18), as per the stance forwarded by Widdowson (1975).

1.5 The structure of the thesis

This study is presented in seven chapters, with their functions described as follows. Chapter 1 presents an introduction to the study. Chapter 2 defines the context of the thesis by addressing the differences between simplified and authentic texts, while offering several statistical analyses to establish statistically the criterion of each text type (i.e. simplified and authentic). Through establishing the difference between simplified and authentic texts, I identify the gap that non-native readers encounter during the transition from simplified texts (despite being categorised as advanced) to their authentic counterparts. Chapter 3 addresses the concepts related to corpora and stylistics: corpus stylistics and lexical style. This chapter highlights linguistic insights which can be provided via the stylistic analysis of literary texts by means of corpus stylistics. Chapter 4 focuses on a detailed description of Dickens's Complete Corpus (the DCC) and how it has been constructed. This chapter describes the stages of determining Dickens's works, and the manner in which the DCC has been compiled. The criteria for constructing the corpus are considered in general, and particularly those that relate to the specialised corpus (e.g. size, representativeness). Chapter 5 explores the applications of corpus linguistics, such as the notions of word lists and collocations, and considers how such concepts can form the basis for stylistic analysis that contributes towards the awareness raising of learners and readers of Dickens in respect to the potential stylistic effects that may arise from such linguistic aspects. Chapter 6 is concerned with the analysis of the DCC, demonstrating the preliminary results of analysing Dickens's complete works in terms of the number of tokens, the word list, and keywords. This chapter also addresses the potential for those wordlists to

illuminate the DCC in terms of the nature of Dickens's works, and what stands out as distinguished in his works, by identifying 'domain-specific linguistic patterns' (Bowker & Pearson 2002: 4). The pedagogical relevance of the most frequently occurring words in the DCC is also discussed, along with strategies for EFL learners to utilise this resource when reading Dickens's works. Furthermore, this chapter establishes the value of specialised corpora like the DCC in facilitating the reader's engagement with that given area, such as Dickens's works (O'Keeffe, McCarthy & Carter 2007: xiii). Chapter 6 also features several stylistic insights into Dickens's works, focusing on his choice of lexicon and how that serves to create a specific impact on the reader. Chapter 7 concludes the study by confirming the contributions achieved in this thesis, and then considering the possible directions where such research could be advanced in future. Finally, a number of appendices related to the DCC are included.

It is hoped that this study contributes towards illustrating a clear and practical example of how corpus stylistics can successfully illuminate interesting characteristics and facets in the context of reading and teaching literary texts, and in particular Dickens's works as one of the commonly used literary texts employed in the teaching of the English language. This study aims to provide a model for the teaching of a literary text of Dickens, whereby the readers acquire the lexicon necessary to comprehend the text and achieve the awareness of his lexical choice in order to engage with and appreciate Dickens's style. The potential of corpus stylistics, as discussed and with reference to Dickens's lexicon and non-native readers, will hopefully pave the way to tackling authentic literary texts, namely, illuminating how the acquisition of new lexicons can be developed for pedagogical purposes based on a rigid approach with stylistic insights. It is anticipated that through this study the teaching and learning of new lexicon will be further extended from the conventional practice found in pedagogical studies, towards a more objective approach grounded in evidence from corpus stylistics. It is also hoped that this study provides insights into literary criticism, to enable a more scientific-oriented and rigorous approach for the evaluation and

appreciation of the literary texts, beyond the domain of subjectivity and self-impressionism. This study in essence is intended to promote ‘the use of corpora as tools in the hands of teachers and learners [which] is moving on very hesitatingly’ (Gavioli 2005: 1) with reference to vocabulary learning and teaching, and which has received scant attention in the literature of SLA, as observed in detail by (Milton 2009: 1–5).

1.6 Conclusion

In this chapter I have argued that those learners who are at the advanced levels and aspire to progress forward towards reading authentic texts require systematic assistance, in which the task of reading itself is simplified, as opposed to simplifying the reading materials themselves. I have also highlighted the value of learning vocabulary from within their natural discourse, while simultaneously acknowledging the usefulness of engaging with specifically designed vocabulary lists such as the Dickens’s World List.

Behnam (1996) reminds the reader that stylistics offers a direction by which valuable insights can be realised regarding literary discourse and through which it is possible to form concluding responses, although ‘readers, students, teachers and analysts of literature should bear in mind the limitations and problems of stylistics before they resort to it as an analytic tool’ (Behnam 1996: 27); however, Behnam (1996) holds firm in the belief that ‘flexible pedagogical stylistics can do a great deal’ for those working and studying in the field of SLA (Behnam 1996: 27). In conclusion of the process thus far, I have attempted to illuminate the role of stylistics in the study of literature, exploring some of the stylistic devices, that is, lexical items and collocations. Ultimately, via the employment of a straightforward but rigorous and efficient approach to the works of Dickens, and grounded in the investigation of lexical items, this study aims to emphasise the significance of recognising the concept of style in order to ensure its accessibility to language learners placed at the advanced level. It is thus

believed that the stylistic approach will facilitate learners' ability to explore and appreciate the manner in which the language of Dickens functions.

Chapter 2 The Nature of Authentic and Simplified Texts

2.1 Introduction

This chapter presents the context in which this study is focused. I explain the context within which learners of **EFL** in the *English Language and Translation Department of Qassim University, Saudi Arabia*, are taught simplified texts during the three years of their BA programme, before the transition in the fourth and final year (levels 7 & 8) to reading authentic texts. The learners commence this shift by engaging with authentic literary texts (as defined in Section 2.3) including plays such as Shakespeare's *Hamlet* and novels such as Dickens's *Oliver Twist* and *A Tale of Two Cities*. The move from simplified texts to their more authentic counterparts represents a challenge for the learners as the linguistic features in text types are clearly distinct, as will be demonstrated in Section 2.4 when the two types of texts are compared. It is highly likely that learners of English will encounter a number of challenges when engaging with authentic texts, as Gardner (2007) suggests that readers of such texts may be deprived of comprehension due to the lexical demands necessary to reach a level where the learning process can take place. The aim of this research, as stated in Chapter 1, is to support learners and facilitate their transition from reading simplified texts to a level where they can cope, comprehend and appreciate Dickens's authentic texts (i.e. unsimplified versions of his works). This chapter will discuss the concepts of *authentic texts* and *simplified texts* and the related linguistic features presented in such text types through a comparative analysis of the two different types of texts which can be found in Dicken's novels in their respective authentic and simplified forms. I also address the pedagogical applications of the use of authentic texts in the context of those learners of English deemed to be at the advanced level.

I consider that utilising the simplified versions of authentic texts for pedagogical purposes is justified, provided that they satisfy the needs of the non-native learners at the earlier stages of their foreign/second language acquisition. However, this use of simplified texts will inevitably lead to the acquiring of an isolated language style, since it has not manifested naturally as a communicative exchange amongst the native speakers of the target language and is thus unlikely to result in effective potential for long-term communication between native and non-native speakers; the simplified language being engineered for teaching purposes and thus artificial texts, rather than emerging from native speakers' discourse. I take the position that the ultimate aim of learning and teaching a foreign/second language should be to establish proper communication between the learners and the society in which the target language is found, English speaking communities in this context (see Widdowson 1978). Due to the fact that simplified texts lack the innate nature of being a natural and authentic (or genuine, as Widdowson (1978: 88–9) terms it) means of communication between the learners and native users of the target language, they are not effective at advanced levels where learners are expected to transition towards proficiency and becoming competent users of the target language, where they can display a good command of communicative competence. In order to achieve this required level of establishing a proper mode of communication between the learners and the native speakers of the language in the general sense, it is believed that non-native learners should be exposed to authentic materials as soon as they have reached the level at which they can cope with such authentic materials.

In relation to the communicative competence discussed in Chapter 1, I believe that advanced learners ought to commence reading (or listening to) authentic materials when reaching their *advanced levels*. **Through exposing advanced learners to more authentic materials they can learn how to enhance their communication in the target language, since such authentic materials may reflect that which they will encounter, and thus need to accommodate in the native speaking environment.**

Dickens's works in this sense reflect what learners may encounter in the process of studying the English language in many tertiary educational institutions. With regards to advanced learners reading Dickens's works, the task of reading itself becomes a means for the acquisition of new vocabulary and the observation of grammar at play, while simultaneously developing a deeper understanding of the English language. Since such learners will eventually encounter authentic texts, authentic materials offer invaluable access to the target language, which alongside the target culture represent an important component of communicative competence that extends beyond the knowledge of linguistic structures to the domains of nuanced usage and effective perception. Hall (2015) confirms that the intrinsic 'communicative competence' component of all languages entails the use, understanding and collaborative 'play with puns, proverbs and the sounds and shapes of words' (Hall 2015: 28). The level of communication achievable by advanced learners through engaging with Dickens's works thus extends beyond merely using language purely as a communicative means, through opening windows that allow learners from different cultures 'to cope with both the subject matter and skills associated with that discipline' (Walsh 2006: 7).

While the notion of simplified materials can be useful at the beginning levels, as mentioned above, learners should avoid becoming reliant on any type of modified materials; otherwise, they may be misinformed by the input from such learning materials and will subsequently be unable to cope with the type of communication that they are likely to face in the real world. Crossley et al. (2007), quoting Davies and Widdowson (1974), highlight that '[o]ne shortcoming of simplified texts ... is that lexical simplification may lead to a reliance on more common words in the text' (Davies & Widdowson 1974, cited in Crossley et al. 2007: 20). Moreover, Auerbach and Burgess (1985) consider that simplification fails to present 'situationally realistic content', features 'oversimplified reality' and 'often lead[s] to communicatively unrealistic passages' (Auerbach & Burgess 1985: 478, 483, 488, respectively), while

Crossley et al. (2007: 16–7) report some useful criticism on the genre of simplified language.

A further aspect to be taken into consideration and which is a **quality of** authentic texts is that they comprise key stylistic features of the authentic language, features which might actually facilitate comprehension and interpretation. Widdowson (1978) considers that simplification as ‘a process of lexical and syntactic substitution’ is ‘a kind of translation’, emphasising that the issues with such a process ‘can often result in a distortion of use’ (Widdowson 1978: 88). Widdowson (1978) proceeds to confirm that ‘in trying to make meanings explicit with a restricted range of usage, [this] alters the relative prominence of the elements in the original proposition and so changes their function’, resulting in it being ‘almost impossible to avoid distortion’ (Widdowson 1978: 89). Moreover, Hyland (2007) finds that the ‘commitment to exploiting relevant and authentic texts in the classroom’ increases ‘awareness’ of the ‘purpose and ... linguistic and rhetorical features’ of these authentic texts (Hyland 2007: 397). Such tasks will, as Breen (1985) suggests, ‘develop an authentic interpretation’ (Breen 1985: 69, note 6; emphasis in original).

2.2 Simplified texts

Simplified texts (also referred to as *artificial* or *modified* texts; see Flowerdew 2013: 174 and Crossley et al. 2007: 25, respectively) are those materials which have been adapted, written or rewritten for pedagogical purposes to facilitate comprehension. According to Simensen (1987), simplified texts are controlled and interpreted in three principles or categories: the ‘control of information, control of language and control of discourse and texts structure’ (Simensen 1987: 45). Moreover, Crossley et al. (2007) elaborate on the nature of simplified texts by underscoring that

simplified texts are texts written (a) to illustrate a specific language feature such as the use of modals or the third-person singular verb form; (b) to modify the amount of new lexical input introduced to learners; or (c) to control for propositional input, or a combination thereof.

(Crossley et al. 2007: 16)

Waters (2009: 317–8) explains that texts employed for pedagogical purposes need not necessarily remain authentic at all times in order to scaffold the learning process, as simplified texts in textbooks, for instance, may enhance the learner’s ability to comprehend and engage with the target language. On some occasions, fulfilling the needs of the learners makes it prudent to select simplified texts over their authentic counterparts, particularly in the early stages of language development, **as** in the simplified text the level of lexicon is moderated to meet the level of the reader. Gardner (2007) indicates that the ability to control vocabulary ‘seem[s] possible through materials and communicative contexts that have been linguistically engineered to control for vocabulary presentation’ (Gardner 2007: 248).

Crossley, Yang and McNamara (2014) assert that the intention of the simplification of texts is to enhance their readability, and thus their accessibility to learners, with the modifications generally being applied at the lexical and syntactic levels. Simplifying text can be achieved ‘through a process of elaboration, which clarifies message content and structure through [the] repetition of key ideas and the paraphrasing of difficult terms’ (Crossley, Yang & McNamara 2014: 92). It is worth underscoring here that simplified **literary** texts are employed as a vehicle for language input within the context of second or foreign language learners only at the elementary and intermediate levels (Crossley et al. 2007).

Amongst the characteristics of simplified texts is the length, which tends to be far shorter than that of their authentic counterparts. Table 2.1 presents data regarding the length of two of Dickens’s novels, which underscores a significant difference (see Section 2.4 below). The grading of the text typically results in a reduction in the *amount of information* conveyed, thus rendering simplified texts easier for the learners to comprehend. Furthermore, the *reading process* itself is assisted as control over the variables such as rhetorical organisation, length, syntactic complexity and lexical richness is introduced, as these have been found to be significant factors affecting the reading process (see Carrell 1991). Aluísio et al.

(2008) also **highlight** that simplified texts tend to have 'limited vocabulary' and 'restricted sentence structure' (Aluísio et al. 2008: 16), which makes it more feasible for non-native readers to engage with and comprehend them.

Despite the texts being easier to understand, readers of simplified **literary** texts are deprived of the authentic response, with Widdowson (1979) suggesting that since such **editions** do not reflect 'natural language behaviour', the response itself will be inauthentic (Widdowson 1979: 161). Nevertheless, failure to deal with authentic materials which were produced with 'an authentic communicative objective in mind' (Swaffar 1985: 17) leads to an inadequate form of communication. Moreover, simplified **literary** texts do not provide richness in respect to the 'amount of semantic detail' (Gass 2003: 231), which deprives learners of the ability to practise the skills necessary to comprehend authentic texts. And lastly, the simplification of the texts 'often produces unnatural target language models' due to the '[r]emoval of unknown linguistic forms [which] inevitably denies learners access to the very items they need to learn' (Long & Ross 1993: 29–30).

Johnson and Johnson (1998) suggest that there are three directions through which to approach the language difficulties encountered in texts: extending learners' vocabulary range in order to ameliorate the difficulties of engaging with authentic texts; employing simplified texts by 'select[ing] more linguistically accessible text' which is 'controlled in terms of syntax (from simple to complex structures) and lexis (from frequent to less frequent words)'; and 'teach[ing] strategies to cope with unknown language' such as by 'guess[ing] the meaning from the form of the word' or 'ignoring the word and carrying on' (Johnson & Johnson 1998: 332). Since my study addresses a specific situation where learners are reading Dickens's works as authentic literary prose, the intended aim is to elevate the learners' awareness levels in terms of lexis and style to meet the level of authentic works such as Dickens's, as will be detailed in Chapter 5.

This naturally leads on to consideration of the relevance and importance of authentic literary texts such as Dickens's to facilitate learners of English in their comprehension of the author's prose, as introduced in Section 2.3.

2.3 *Authentic texts*

Dickens's original works can be classified as authentic texts, as according to Leaver, Ehrman and Shekhtman (2005) authentic texts are 'those writings, broadcasts, or speech events prepared by native speakers for native speakers' (Leaver, Ehrman & Shekhtman 2005: 20). Johnson and Johnson (1998) consider authenticity as 'genuine instances of language use as opposed to exemplars devised specifically for language teaching purposes' (Johnson & Johnson 1998: 24), while Singleton (2000) emphasises the sense that authentic texts in a given language (either written or spoken) are created with native speakers in mind. This excludes materials (written or spoken) developed for educational purposes, even if they are produced by native speakers of the target language. Rixon (2000) suggests that the social dimension of the authenticity indicates towards the fact that authentic text 'was first conceived as a way of communicating amongst **NATIVE SPEAKERS** of a particular language, with no intention on the part of its originators for it to be used as an instrument for teaching that language to learners' (Rixon 2000: 68; bold and small capitals in original). Furthermore, Seidlhofer (2002) underscores the social role of such produced texts as they 'occur naturally in native speaker communication' and 'have not been designed specially for language teaching purposes' (Seidlhofer 2002: 220). Therefore, the term *authentic texts* refers to those texts (including spoken) produced by native speakers for an audience comprising of their peers in order to fulfil social communication, as opposed to being created for teaching purposes or with second/foreign language learners in mind (Leaver & Shekhtman 2002).

Concerning the characteristics of authentic texts, in comparison with their simplified counterparts the former are likely to have some features that differentiate

them from simplified or modified texts. These characteristics include, for instance, their significant complexity (Stephenson 2009), which can be clearly evident in the structure of the texts (syntax) and the lexical level (semantics). They are considered to be challenging due to the 'quantity of unfamiliar language', which 'is often seen as a measure of difficulty' (Rixon 2000: 68). Gascoigne (2008) considers that authentic texts tend to be lengthy, that they 'often have a natural redundancy of vocabulary' and a characteristic *discourse structure*, besides conveying a 'theme that is useful to the L2 reader' (Gascoigne 2008: 78). Fidler (2003) concurs with the theme aspect, while finding authentic texts 'intellectually stimulating' (Fidler 2003: 113). An additional characteristic of authentic texts is related to their representativeness, as according to Mauranen (2004) they 'represent the target language' (Mauranen 2004: 103). Furthermore, Maley and Duff (1990) point out that the authenticity of literary texts 'offer[s] genuine samples of a very wide range of *styles, registers and text-types* at many levels of difficulty' (Maley & Duff 1990, cited in Gilroy & Parkinson 1996: 215; emphasis added).

Having defined and identified the characteristics of authentic texts, it is now pertinent to indicate towards the advantages that learners can realise through engaging with them. Johnson and Johnson (1998) suggest that there are two criteria which justify why authenticity is important. Firstly, it exposes learners to a language level which echoes that which native speakers engage with, which is considered 'necessary' for the learners in order to 'properly ... interpret texts' (Johnson & Johnson 1998: 24). Secondly, authentic texts can represent 'a means of communicating content' rather than creating artificial texts to fulfil this social task (Johnson & Johnson 1998: 24). By analysing a range of studies, Bernhardt (2011) affirms 'the significance of authenticity in the reading process and that research should be directed toward understanding second-language reading within the context of authentic texts that are neither manipulated nor overanalyzed' (Bernhardt 2011: 61). Moreover, Moyer (2009) suggests that authentic input scaffolds linguistic

fluency (see Moyer 2009: 161). Alluding to what he refers to as ‘communicative needs’, Singleton (2000) affirms that the linguistic characteristics that learners need to understand and express can be attained by exposing them to ‘authentic samples of the target language’, which then leads on to ‘activities associated with real-life needs’ (Singleton 2000: 216). Furthermore, Rixon (2000) indicates towards a prominent feature of authentic materials, namely the *quality* of the language data employed to promote the communicative competency of the learners.

2.4 Comparative analysis of authentic and simplified texts

In this section, I will illustrate the differences between authentic and simplified texts. The purpose of this comparative analysis is to identify the significant contrasts between the two types of texts and to underscore that learners require assistance in order to tackle this transference. That is, learners need help to make the successful transfer from reading simplified texts (although they are categorised as advanced level) to reading the authentic counterparts, bearing in mind the complexity that accompanies this shift in terms of the lexical and structural levels. For non-native readers, the transition from simplified texts (albeit being classified as *advanced*, addressing the level of the readers) to authentic texts represents a significant challenge as there is a noteworthy difference between simplified and authentic texts in respect to the texts’ length, the lexical sophistication (i.e. the number of headwords) and the syntactic level. This comparison is not to argue for or against the use of authentic or simplified texts (e.g. see Aghagolzadeh & Tajabadi 2012; Crossley et al. 2007), but rather to establish the contexts addressed in this study. Moreover, despite the modifications of (authentic) texts to transform them into simplified versions generally occurring at the lexical and syntactic levels, this study’s focus is on the lexical boundaries and the stylistic effects associated with them, as will be discussed in Chapter 5.

The comparison will be conducted on two of Dickens's novels, *Oliver Twist* and *Great Expectations*, in both their authentic and simplified versions and from two different publishers: Oxford University Press (*Oxford Bookworms Library*) and Pearson Education (*Penguin Readers*). *Oxford Bookworms Library* and *Penguin Readers* are both series of carefully controlled graded texts that present, amongst others, simplified versions of Dickens's *Oliver Twist* (Oxford University Press) and *Great Expectations* (Penguin Books Ltd), as shown in Table 2.1.

Table 2.1 The simplified versions of two of Dickens's novels

Simplified versions	Oliver Twist	Great Expectations
Publisher	Oxford University Press	Pearson Education
Series	<i>Oxford Bookworms Library</i>	<i>Penguin Readers</i>
Level	Stage 6 – advanced	Level 6 – advanced
Headwords	2,500	3,000
Word count	26,560	35,437

Prior to conducting the comparative analysis between the selected Dickens's texts in their two different types (i.e. simplified and authentic), a number of general differences that are identified at the initial stage will be introduced. Using AntConc 3.4.4w, 'a freeware, multi-platform, multi-purpose corpus analysis toolkit' (Anthony 2005: 729), the two versions of each work (the simplified and the authentic) were compared in terms of the number of *headwords* and the *total number of running words (tokens)*, with the comparison revealing the following data presented in Table 2.2.

Table 2.2 Comparison between the simplified and authentic texts of two of Dickens's novels

Text	Simplified text		Authentic text		Headword increase	
	<i>word tokens</i>	<i>headwords</i>	<i>word tokens</i>	<i>headwords</i>		
Oliver Twist	26,560	2,500	165,925	7,397	4,397	195.88%
Great Expectations	35,437	3,000	188,900	7,858	4,858	161.93%

Table 2.2 illustrates two factors that impact on the learners' ability to read and comprehend Dickens's works: the text's length and the number of headwords, which represent the vocabulary diversity present in each text. In terms of length, the authentic text of *Oliver Twist* was found to contain 165,925 words (tokens), as opposed to the 26,560 that feature in the simplified version for advanced levels, with the former thus being more than six times longer. Staying with the same novel, the intensity of headwords (lexemes) in the authentic text was close to three times the size when compared with the simplified text, increasing from 2,500 headwords in the simplified text to 7,397 in the authentic text; the percentage increase in headwords from the simplified to the authentic text being 195.88% in total. The principle is virtually identical for *Great Expectations*, if not more so. The word count of the simplified version of *Great Expectations* is 35,437, while the authentic version is 188,900, with the latter being an increase to more than five times longer; that is, a percentage increase of 433.06%. Concerning the load (i.e. the number of headwords in relation to the running words of the texts; tokens) of the headwords in *Great Expectations*, it increases from 3,000 in the simplified version to 7,858 headwords in the authentic, representing an increase of more than two and half times (161.93%).

To place this increase in the number of headwords from the advanced level to its authentic counterpart into context, the percentage increase in headwords will be presented between the levels of the simplified texts themselves in the two series, *Oxford Bookworms Library* and *Penguin Readers*, and then a comparison made with the shift from the (advanced) simplified texts to the authentic versions, as presented in Table 2.3.

Table 2.3 Headwords' percentage increase between each stage/level at the simplified levels, and at the authentic text level

Stage/Level		<i>Oxford Bookworms Library</i>		<i>Penguin Readers</i>	
		Oliver Twist		Great Expectations	
0	Starter / Easystart	250		200	
1	Stage 1 / Level 1	400	60%	300	50%

2	Stage 2 / Level 2	700	75%	600	100%
3	Stage 3 / Level 3	1,000	42.9%	1,200	100%
4	Stage 4 / Level 4	1,400	40%	1,700	41.7%
5	Stage 5 / Level 5	1,800	28.8%	2,300	35.3%
6	Stage 6 / Level 6	2,500	38.9%	3,000	30.4%
Average increase		47.6%		59.57%	
Authentic texts		195.88%		161.93%	

In both the *Oxford Bookworms Library* and *Penguin Readers* series the increase of headwords shifts between increments of 29% and 100%, depending on the level and series. Placing these idiosyncrasies aside, it is important to note the increase in headwords from the advanced to authentic levels, where there is a significant increase in both series, underscoring the breadth of the transition and challenges for learners attempting to engage with authentic texts.

Two observations should be made here. Firstly, the percentage increase in headwords typically follows a downward trend when progressing through the reading scale, and therefore the norm when shifting from the advanced level to the authentic text might be expected to follow this pattern; however, what has been revealed is quite the opposite, essentially a sudden and significant increase in the headwords, although there is some variation in the actual increase in headwords from the advanced to authentic texts, as shown in Table 2.2, where the increase in *Oliver Twist* was 195.88% and in *Great Expectation* was somewhat lower at 161.93%.

The second observation is that the percentage increase in both the series within the simplified levels/stages does not exceed 100% at its maximum, occurring in *Penguin Readers* twice when moving from *Level 1* to *Level 2* and from *Level 2* to *Level 3*. However, by considering the increase in the number of new headwords when moving from the advanced simplified text to the authentic, it can be noticed that this is the highest increase in terms of percentage, and conflicts with the norm where the percentage of headwords should be declining when readers move to the upper levels. Although the shift from stage/level 6 to the authentic texts is not

strictly part of the series, **and** so it could be argued to fall outside the trends described in Table 2.3, that is, what non-native learners face when they move to the authentic texts **in the real situation described above**. This rise in the number of headwords clearly underscores the gap between the two levels, that is, the so-called *advanced* and *authentic*, and thus exemplifies that learners require additional assistance in successfully navigating the transition from advanced levels to dealing with authentic texts. Therefore, traversing this shift cannot be claimed to be a task easily achieved by merely reading and rereading the literary work, as is sometimes suggested in the traditional approach to teaching literature.

The two primary linguistic features that are typically discussed and presented over others are lexical density and the complexity of the grammatical structure of texts. Crossley and his colleagues (Crossley et al. 2007; Crossley, Allen & McNamara 2012; Crossley, Yang & McNamara 2014) have conducted a range of studies investigating the differences between simplified and authentic texts. What is interesting about their pioneering research is that it compares the two types of texts (i.e. simplified and authentic) in terms of their linguistic features and is conducted *statistically*. A further interesting facet of their work is that they also carry out comparisons between the simplified texts themselves, for instance simplified texts for *beginners*, and simplified texts for *intermediate* or *advanced* levels, and consequently identify several differences in their *linguistic* features.

Adhering to the purpose of this study, focus will be placed on the lexical domain, and the question that has been addressed in part above (What is the difference between simplified and authentic texts in terms of their lexical richness?), while Chapter 3 will address the question of how non-native readers can overcome the **significant increase** in headwords identified in this chapter between simplified and authentic texts in terms of their lexical aspects.

2.4.1 The lexical profile statistics of simplified and authentic samples

It may prove beneficial to begin with examples of the opening paraphrases in both the simplified and authentic versions of *Oliver Twist* and *Great Expectations*. These examples will be restricted to the first paragraph in the simplified versions, but then the contents will need to be matched with that of the authentic counterparts. For this comparison, the four quoted passages will be compared using AntWordProfiler 1.4.0w (Anthony 2013) after saving the four passages in four individual .TXT files (The significance of the coloured text in the quoted passages will be explained below).

Oliver Twist

The first paragraph of the simplified version of *Oliver Twist* reads:

Oliver Twist was born in a workhouse, and when he arrived in this hard world, it was very doubtful whether he would live beyond the first three minutes. He lay on a hard little bed and struggled to start breathing.

(*Oliver Twist*, Oxford Bookworms Library, Stage 6, Chap. 1)

Table 2.4 *Oliver Twist* – lexical profile statistics for simplified paragraph

File	Token	Type	Group	Token%
basewrd1.txt	35	28	28	87.5%
basewrd2.txt	3	3	3	7.5%
basewrd10.txt	1	1	1	2.5%
basewrd31.txt	1	1	1	2.5%
Total	40	33	33	100%

The respective authentic text reads:

Among other public buildings in a certain town, which for many reasons it will be prudent to refrain from mentioning, and to which I will assign no fictitious name, there is one anciently common to most towns, great or small: to wit, a workhouse; and in this workhouse was born; on a day and date which I need not trouble myself to repeat, inasmuch as it can be of no possible consequence to the reader, in this stage of the business at all events; the item of mortality whose name is prefixed to the head of this chapter.

For a long time after it was ushered into this world of sorrow and trouble, by the parish surgeon, it remained a matter of considerable doubt whether the

child would **survive** to bear any name at all; in which case it is **somewhat** more than probable that these **memoirs** would never have appeared; or, if they had, that being **comprised** within a couple of pages, they would have **possessed** the **inestimable merit** of being the most **concise** and **faithful specimen** of **biography**, **extant** in the **literature** of any age or country.

(*Oliver Twist*, Oxford World's Classics, Chap. 1)

Table 2.5 *Oliver Twist* – lexical profile statistics for authentic paragraph

File	Token	Type	Group	Token%
basewrd1.txt	153	83	76	80.95%
basewrd2.txt	10	10	10	5.29%
basewrd3.txt	8	8	8	4.23%
basewrd4.txt	5	5	5	2.65%
basewrd5.txt	4	4	4	2.12%
basewrd6.txt	1	1	1	0.53%
basewrd8.txt	1	1	1	0.53%
basewrd9.txt	2	2	2	1.06%
basewrd10.txt	2	1	1	1.06%
basewrd11.txt	1	1	1	0.53%
basewrd13.txt	1	1	1	0.53%
basewrd16.txt	1	1	1	0.53%
Total	189	118	111	100%

Great Expectations

The first paragraph of the simplified version of *Great Expectations* reads:

My father's family name being Pirrip, and my **Christian** name **Philip**, my **infant tongue** could make of both names nothing longer than **Pip**. So I called myself **Pip**, and came to be called **Pip**. Having lost both my parents in my **infancy**, I was brought up by my sister, Mrs **Joe** Gargery, who married the local **blacksmith**.

(*Great Expectations*, Penguin Readers, Level 6, Chap. 1)

Table 2.6 *Great Expectations* – lexical profile statistics for simplified paragraph

File	Token	Type	Group	Token%
basewrd1.txt	46	36	30	79.31%
basewrd2.txt	1	1	1	1.72%
basewrd3.txt	2	2	1	3.45%
basewrd8.txt	1	1	1	1.72%
basewrd11.txt	3	1	1	5.17%

basewrd31.txt	3	3	3	5.17%
Non-level list words	2	2	2	3.45%
Total	58	46	39	100%

And the respective authentic text reads:

My father's family name being Pirrip, and my Christian name Philip, my infant tongue could make of both names nothing longer or more explicit than Pip. So, I called myself Pip, and came to be called Pip.

I give Pirrip as my father's family name, on the authority of his tombstone and my sister – Mrs Joe Gargery, who married the blacksmith. As I never saw my father or my mother, and never saw any likeness of either of them (for their days were long before the days of photographs), my first fancies regarding what they were like, were unreasonably derived from their tombstones. The shape of the letters on my father's, gave me an odd idea that he was a square, stout, dark man, with curly black hair. From the character and turn of the inscription, 'Also Georgiana Wife of the Above,' I drew a childish conclusion that my mother was freckled and sickly. To five little stone lozenges, each about a foot and a half long, which were arranged in a neat row beside their grave, and were sacred to the memory of five little brothers of mine – who gave up trying to get a living, exceedingly early in that universal struggle – I am indebted for a belief I religiously entertained that they had all been born on their backs with their hands in their trousers-pockets, and had never taken them out in this state of existence.

(*Great Expectations*, Penguin Classics, Vol. 1, Chap. 1)

Table 2.7 *Great Expectations* – lexical profile statistics for authentic paragraph

File	Token	Type	Group	Token%
basewrd1.txt	200	106	89	82.64%
basewrd2.txt	11	11	11	4.55%
basewrd3.txt	11	11	11	4.55%
basewrd4.txt	2	2	2	0.83%
basewrd6.txt	2	2	2	0.83%
basewrd7.txt	1	1	1	0.41%
basewrd8.txt	2	2	2	0.83%
basewrd11.txt	3	1	1	1.24%
basewrd12.txt	1	1	1	0.41%
basewrd31.txt	4	4	4	1.65%

basewrd33.txt	2	2	1	0.83%
Non-level list words	3	2	2	1.24%
Total	242	145	127	100%

In order to examine the vocabulary load contained within these four short passages, the concept of ‘word family’ requires introduction. In order to understand what *word family* refers to, it is essential to recall the definition of the *headword* as addressed in Chapter 1. The headword, as defined earlier, refers to a word in its basic form that can be extended to other related forms. Bauer and Nation (1993) consider that a word family ‘consists of a base word and all its *derived* and *inflected* forms that can be understood by a learner without having to learn each form separately’ (Bauer & Nation 1993: 253; emphasis added). Therefore, the word family **refers to** ‘a headword, its *inflected* forms, and its closely related *derived* forms’ (Nation 2001: 8; emphasis added). These in sum are three categories, namely, the *headword* (also referred to as the *base*, *root* or *stem* word) and its *inflection* and *derivation*. Schmitt (2000) clarifies what is meant by *inflection* and *derivation* (derivative forms), with the inflected forms of a headword resulting from the inclusion of affixes for grammatical purposes; an example of an inflected word being when the base form (headword) *walk* morphs into *walked*, *walking* and *walks*. These three forms, together with the headword, still function as the same part of speech category, that is, they are all verbs. If adding the affixes to the headword ‘change[s] the word class’, then ‘the result is a *derivative*’ form, such as the derivative forms *stimulative* and *stimulation* originating from the headword *stimulate* (Schmitt 2000: 2) **(Note that while the terms *lexeme*, *headword* and *word family* are used interchangeably in this thesis to refer to the same concept, each is specific to a particular field of study, where the *lexeme* is employed in linguistics (semantics) to describe a certain unit of lexical meaning that exists irrespective of how many inflective endings it may have, or the total number of words that comprise it; the *headword* is found in the field of English as a second language (SLA) to define a word under which a series of**

dictionary entries appear; and the *word family* is used in corpus linguistics to describe a word in its base form, alongside any inflected or derived forms created through affixation).

The difference between the *lemma* and the *headword* (basic form) is that lemmas maintain the same category in respect to the part of speech. Moreover, when discussing the concept of word families, a headword will include both the inflected and derived forms ‘even if the part of speech is not the same’ (Nation 2004: 6; emphasis added), as stated above. The result will then be that a headword, as Nation (2006) suggests, can include more than one lemma. Using the example of the headword *abbreviate*, this takes the following family members: *abbreviate*, *abbreviates*, *abbreviated*, *abbreviating*, *abbreviation*, *abbreviations*. If these were treated as lemmas, *abbreviate* and *abbreviation* would represent two different lemmas, the first of which embraces the four family members of *abbreviate*, *abbreviates*, *abbreviated* and *abbreviating* in its verb class; with the latter taking on the two members of *abbreviation* and *abbreviations* as nouns.

2.4.2 Word-family lists

Now, in order to examine the vocabulary load of these four short passages, Paul Nation’s word-family lists are utilised. These are high frequency wordlists developed from the analysis of both the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). Regarding these lists, and as defined previously, a word family consists of a *headword* (root form) and its closely related inflected and derived forms as the family members of the same headword (Nation 2009: 106). The wordlists that Nation (2012) extracted from both the BNC and COCA (presented in Table 2.8 below) echo the patterns, whereby these lists ‘consist of base words and their closely related inflected and derived forms’ (Nation 2001: 34). These wordlists contain no ‘phrases’, but can contain ‘compound words’ (Nation 2006: 66). Finally, the word-family lists ‘are sequenced largely according to

their range and frequency' (Nation 2006: 88). The range of word family indicates to 'how well that word is distributed throughout the corpus' (Leech, Rayson & Wilson 2001: 10), the BNC and COCA in this regard.

The purpose of employing these lists is to identify how commonly occurring the headwords in the examined passages are, and the manner in which they are dispersed in different levels. The word-family lists prepared for AntWordProfiler 1.4.0w (Anthony 2013), as described by Nation (2012), consist of the following:

Table 2.8 The contents of Paul Nation's BNC/COCA word-family lists

Level list	Word-family lists	Groups (headwords)	Types (family members)	Contents
1	basewrd1.txt	1,000	6,857	Wordlists 1-25 contain word families based on their frequency and range in both the BNC and COCA
2	basewrd2.txt	1,000	6,370	
3	basewrd3.txt	1,000	5,880	
4	basewrd4.txt	1,000	4,865	
5	basewrd5.txt	1,000	4,294	
6	basewrd6.txt	1,000	4,102	
7	basewrd7.txt	1,000	3,679	
8	basewrd8.txt	1,000	3,419	
9	basewrd9.txt	1,000	3,196	
10	basewrd10.txt	1,000	2,982	
11	basewrd11.txt	1,000	2,942	
12	basewrd12.txt	1,000	2,754	
13	basewrd13.txt	1,000	2,415	
14	basewrd14.txt	1,000	2,299	
15	basewrd15.txt	1,000	2,283	
16	basewrd16.txt	1,000	2,086	
17	basewrd17.txt	1,000	2,076	
18	basewrd18.txt	1,000	1,933	
19	basewrd19.txt	1,000	1,872	
20	basewrd20.txt	1,000	1,820	
21	basewrd21.txt	1,000	1,651	
22	basewrd22.txt	1,000	1,539	
23	basewrd23.txt	1,000	1,394	
24	basewrd24.txt	1,000	1,296	
25	basewrd25.txt	1,000	1,675	
26	basewrd26.txt	1	1	Contain one nonsense

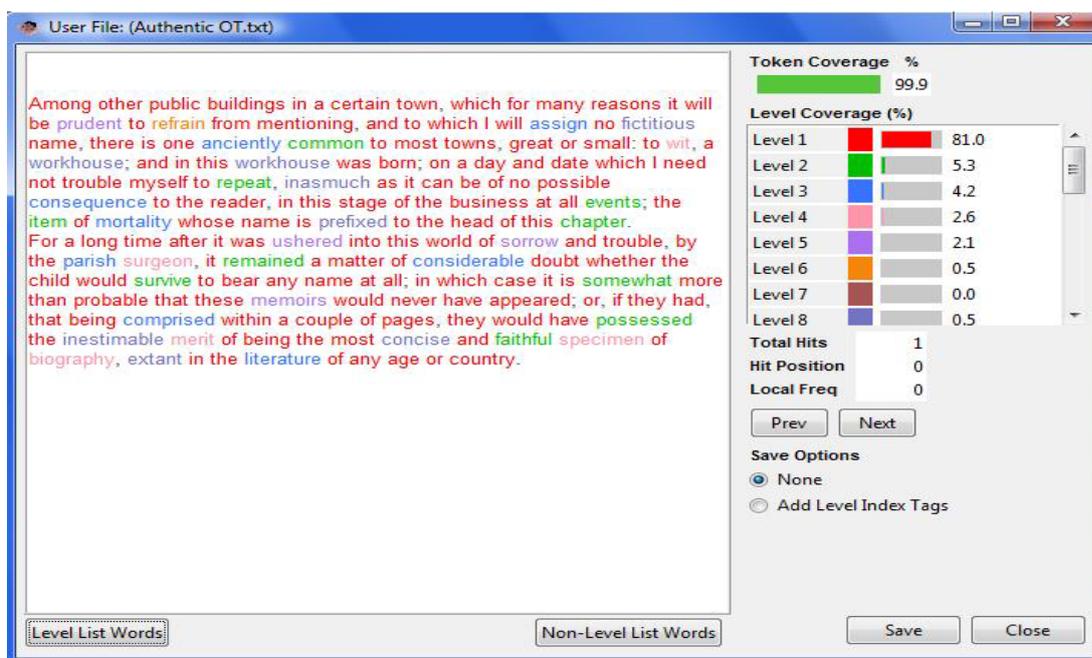
27	basewrd27.txt	1	1	word each
28	basewrd28.txt	1	1	
29	basewrd29.txt	1	1	
30	basewrd30.txt	1	1	
31	basewrd31.txt	21,662	22,409	An ever-growing list of proper names
32	basewrd32.txt	38	196	Marginal words including swear words, exclamations, and letters of the alphabet
33	basewrd33.txt	3,108	6,044	Transparent compounds e.g. <i>lockgates, poolrooms, countertop, duststorm</i>
34	basewrd34.txt	1,083	1,149	Abbreviations

The inclusion of the lists from 26–30, which contain one nonsense word each, are for technical reasons as ‘[t]hey were made to provide space for additional lists and to avoid having to keep changing the names of the proper nouns etc lists’ (Nation 2012: 1). Further details regarding these lists can be found in Nation and Webb (2011: Chap. 8), Nation (2004) and Bauer and Nation (1993).

To conduct this analysis the AntWordProfiler 1.4.0w is utilised, which is described by the writer of the software (Anthony 2013) as ‘a freeware, multiplatform tool for carrying out corpus linguistics research on vocabulary profiling’ (Anthony 2013: §1). It contains two tools. The *Vocabulary Profile Tool* allows the testing of the vocabulary size, besides the generating of extensive word-family lists to provide statistical and frequency information regarding the targeted corpora. It also compares the targeted corpus with ‘a set of vocabulary level lists that can be plain frequency lists of “family lists” based on the research of Paul Nation’ (Anthony 2013: §2). The second tool is the *File Viewer and Editor Tool*, which considers an individual targeted file of a corpus and ‘highlights the different levels of vocabulary [headwords and family members] in the file using a color coding. It also shows the overall coverage of different vocabulary levels’ (Anthony 2013: §2) in the targeted file, that is, how many vocabularies in the analysed file were located in the baseword

level lists, which function like a reference corpus, as seen in Figure 2.1 below. The software features three baseword lists: the first two 1,000 word-family lists from the General Service List (GSL) (1_gsl_1st_1000.txt, 2_gsl_2nd_1000.txt) and the Academic Word List (AWL) (3_awl_570.txt). These baseword lists can be replaced by other wordlists such as ‘family lists’, based on the research of Paul Nation (see Anthony 2013).

Figure 2.1 A view of the authentic passage of *Oliver Twist* in the AntWordProfiler 1.4.0w’s File Viewer and Editor Tool



Using AntWordProfiler 1.4.0w, the four passages were compared against these baseword (word-family) lists in order to identify the number of word families that exist in these passages. Since Nation’s lists are arranged according to their frequency in both the BNC and COCA, this simply means that the higher the group, the less frequent the word; conversely, the more infrequent the family words in the authentic texts are, the more challenging the texts are. This source of difficulty is related to vocabulary sophistication, which is always modified when simplifying texts. Lexical sophistication refers to the ‘percentage of “advanced” words in the text’, while ‘[w]hat is labelled as “advanced” would depend on the researcher’s definition’ (Laufer & Nation 1995: 309). The lexical sophistication is additionally

related to 'the usage of low frequency or more difficult vocabulary' (Malvern et al. 2004: 4). Granger (1998) highlights that the lexical sophistication of texts 'can be obtained by comparing a text against core vocabulary lists to see what words in the text are and are not in the lists' (Granger 1998: 34). I set the first three 1,000 word-family lists as the baseline, after which a headword will be classified in this study as a sophisticated headword, that is, being a rare, less frequent and advanced one. The rationale behind selecting this baseline is that the highest level in the simplified versions was set between 2,550 and 3,000 headwords in both the simplified series (*Oxford Bookworms Library* and *Penguin Readers*), with the assumption that what falls beyond this level is at the upper-advanced or native-like level. Milton (2009) underscores the fact that there 'is no absolute rule as to the point at which a word stops being frequent and becomes infrequent' (Milton 2009: 131). Researchers thus vary in setting the cut-off point for the sophistication of a lexical item ranging from 1,000 word families as in Meara and Bell (2001), to 2,000 word families as in Nation's RANGE software (cited in Milton 2009: 131). Milton (2009) finds that '[t]he popularity of [the] RANGE [programme] suggests that there is something like a consensus emerging at 1000 or 2000 words as a dividing line between frequent and infrequent' (Milton 2009: 131).

Despite the extracted passages being too short to conclude with any general statements about these two studied works of Dickens, they still confirm a point that needs to be raised here, which is related to the coverage of the lexical words in the two different versions (i.e. the simplified and the authentic). As Crossley, Yang and McNamara (2014) confirm, simplified and authentic texts differ 'linguistically at the level of lexical sophistication, syntactic complexity, and cohesion' (Crossley, Yang & McNamara 2014: 93).

By considering the extracts from *Oliver Twist*, we can notice that out of 40 tokens in the simplified version, 35 were found in the first 1,000-word list of Nation's

lists (see Table 2.8), representing 87.5% of the passage. Surprisingly, this means that the reader who knows only the first most frequent word family in Nation's lists can comprehend 87.5% of the text. If we take the second 1,000-word list into consideration then the coverage will increase by 7.5%, covering 95% of the words in the passage. Essentially, the passage is fairly easy to read and understand. Only two tokens were found to be outside of the 2,000 most common word families: the first one is *workhouse*, which appears in the 10th level; and the second is the proper name *Oliver*, which appears in the 31st level.

In contrast to the simplified extract, the authentic passage of *Oliver Twist* has 189 tokens in total, becoming five times longer than the simplified passage. As Table 2.5 reveals, there are 111 word families (groups) as opposed to 33 in the simplified version, thus increasing by approximately three times. Furthermore, the dispersion (see Culpeper 2009; Leech, Rayson & Wilson 2001) of these headwords is scattered over 12 word-family lists (1–6, 8–11, 13 and 16), instead of the four levels found in the simplified version. Through careful analysis of the word *ushered*, which appears in the authentic text of *Oliver Twist* – ‘For a long time after it was ushered into this world of sorrow and trouble’ – two points can be made here. Firstly, this word family – *usher*: *ushered*, *ushering*, *ushers* – appears in baseword5, offering an indication of how infrequently it features in Dickens's works. Secondly, by consulting several of the more popular learner dictionaries, i.e. LDCE, CALD, OALD and CCALD, it was found that none of them list *usher* as an entry (headword). Nevertheless, *usher* does appear in COED and is defined as: ‘verb (i) show or guide somewhere and (ii) (usher something in) cause or mark the start of something new’. This scarce inclusion in these learner dictionaries confirms how infrequently the word is likely to be encountered by English learners.

The simplified passage from *Great Expectations* has 58 tokens dispersed over five baseword lists. Notwithstanding the scarcity of the tokens and headwords (39 compared with the authentic version, as addressed below) two tokens were not

found in any of the twenty-nine 1,000 word-family lists (see Table 2.8). These two tokens are both proper nouns: *Pirrip*, and *Gargery*, the surname of the blacksmith. If the proper nouns *Philip*, *Pip* and *Joe*, which appear in the 31st baseword list, plus the other non-level list tokens *Pirrip* and *Gargery* were excluded (i.e. the proper nouns), the entire passage would be limited to the first three word-family lists (1–3), with the exception of one token that appears in the 8th baseword: *blacksmith*. The result of this exclusion of the proper names from the simplified passage would result in 46 tokens representing 90.2% of the texts and appearing in the first 1,000 word-family list, as shown in Table 2.9 below, thus confirming its simplicity in terms of the number of family types (headwords). Again, this passage would be relatively easy to read and comprehend.

Table 2.9 *Great Expectations* – simplified without proper nouns

Level	File	Token	Type	Group	Token%
1	basewrd1.txt	46	36	30	90.2%
2	basewrd2.txt	1	1	1	1.96%
3	basewrd3.txt	2	2	1	3.92%
8	basewrd8.txt	1	1	1	1.96%
31	basewrd31.txt	1	1	1	1.96%
Total		51	41	34	100%

Regarding the authentic passage of *Great Expectations*, this features 242 tokens, amongst which 127 headwords (groups) can be found dispersed over 12 levels. If the first three levels were considered to be relatively simple, these three baseword word-family lists would cover 91.74% of the authentic text, leaving 8.26% of the text as unfamiliar words and resulting in a challenge in respect to word familiarity. According to Nation (2006: 59) at least 98% coverage of a text is required in order to achieve unassisted comprehension, which **underlines** the source of difficulty in this passage.

This is a comparison of a specific linguistically related feature (lexicon), which obviously fails to capture the entirety of the dimension to which simplified and authentic texts are related (Crossley, Allen & McNamara 2012). The aforementioned findings regarding the authentic passages are consistent with those of Nation (2006), in which he suggests that an 8,000 to 9,000 word-family vocabulary is needed for the comprehension of unsimplified written texts. That is because in the authentic texts, the lexical items reach high word-family levels which are less common than the first, say, two or three 1,000 word-family lists. Hill (1997) asserts:

The gap between the top stage of a graded reader series and unsimplified books remains unsatisfactorily wide. ... The width of the gap arises not from complex syntax, for that is already allowed in the top stages, but from the much greater length ..., the much wider vocabulary ..., and the idiomatic, colloquial, and elliptical use of language in which irony and cultural reference are only two of the barriers to comprehension.

(Hill 1997: 63)

In their analysis of the linguistic characteristics between different levels (*beginners, intermediate and advanced*) in the simplified texts, Crossley, Allen and McNamara's (2012) main findings reveal 'significant differences between levels' in terms of 'lexical sophistication, syntactic complexity, and cohesion' (Crossley, Allen & McNamara 2012: 89). In concluding that the simplified texts are 'less complex lexically', they find that such texts 'contain lower lexical diversity, more frequent words, more familiar words, and more concrete words' (Crossley, Yang & McNamara 2014: 95).

The analysis of the samples from *Oliver Twist* and *Great Expectations* was conducted in this manner in order to identify the assistance required by non-native readers when approaching authentic texts. Nation (2004) affirms that creating wordlists in the context of language teaching is motivated by either the need for 'designing syllabuses' or for 'determining necessities (what needs to be learned)' (Nation 2004: 3), which is conducted to specify the needs of non-native learners.

Following this comparative analysis of simplified and authentic texts, the research question that this study seeks to answer is as follows:

What are the lexemes (headwords) in Dickens's texts that need to be identified in order to assist non-native readers in comprehending and appreciating Dickens's authentic texts?

A detailed analysis of the headwords necessary to engage with Dickens's work will be conducted in Chapter 5 in order to respond to this question that focuses on the analysis of Dickens's lexicon to attain comprehensibility (see Crossley, Yang and McNamara 2014: 93). Assisting learners in their acquisition of Dickens's lexicon will allow them to interact and respond effectively to his work through Widdowson's (1979) emphasis on the 'communicative goals in teaching', which can be achieved by extending the learners' awareness to 'ensure the necessary appropriacy of response' (Widdowson 1979: 162) towards the studied materials.

2.5 Lexicon or structure: prioritising the focus

After presenting the contexts of two sampled works of Dickens as teaching materials, this investigation will now discuss the non-native learners' needs when engaging with such authentic texts. Referring to a specific context where advanced learners are exposed to the authentic texts of Dickens, it is legitimate to consider whether advanced learners of English need to improve their lexicon or awareness of sentence structures prior to reading Dickens's works. Essentially, is it sufficient for learners to understand the structure of English (i.e. the grammatical rules) in order to be able to read and understand, and even appreciate Dickens? Toolan (2011), for instance, in his analysis of the Prelude of George Eliot's *Middlemarch*, argues that 'cop[ing] with sentences that are quite complexly structured' (Toolan 2011: 181–2) facilitates fluent reading without the need to fill any background knowledge gaps.

Taking other related factors into consideration should lead to the more reasonable decision to either focus on the structure of the text or the lexicon,

dependent upon the context and the needs of the learners. Toolan (2011) was perhaps assuming or addressing those readers who are native speakers of English; otherwise, it would be clear that his proposal for the comprehension of complexly structured text fails to consider at least two main factors concerning non-native readers: their vocabulary range (notwithstanding the necessary pre-existing knowledge, placed to one side for the present), and that non-native speakers approach and process reading in a slightly different manner from native speakers, as will be explained below. In terms of vocabulary size, '[t]he lexicon of an average native speaker of English contains about 30,000 words' (Radford et al. 2009: 199), compared with 2,000–5,000 words for non-native speakers (Kelly 1991, cited in Olejniczak 2006: 269), with another higher estimation being approximately 10,000 words (Kaufmann 2003: 119). There are additional estimates that focus on the number of word families known, approximating 17,000 base words (headwords) for educated native speakers of English (Goulden, Nation & Read 1990: 341; see also Daller, Milton & Treffers-Daller 2007).

It would not be reasonable to query in this context whether advanced learners should focus primarily on the lexicon or the complex structure of the authentic texts, as such a question assumes a contradiction between the employment of vocabulary knowledge and grammar rules which creates a false dichotomy. Raising such a question leaves the learners isolated between either prioritising grammatical rules or vocabulary, which is an oversimplification of a complex issue, that is, reading in a foreign language such as reading Dickens's works as authentic texts. Asserting only two choices about such a complex challenge that is affected by a range of factors is simply an error, a '*false dilemma fallacy*' (Browne & Keeley 2007: 91), *since the either-or fallacy assumes only two alternatives, while there are many more facets involved in the process of reading and comprehending a text*. Rather, the question should address a specific context such as where non-native readers (classified at the advanced level) are exposed to authentic texts which are considered as being

relatively complex texts, in terms of their structure and vocabulary diversity (number of variant headwords, not their inflected forms). The question then should be: Where learners of English are deemed as being at an advanced level, should greater focus be placed on the structure of the authentic texts or the lexicon? I argue for my position below through addressing what should matter to learners at the advanced level. By prioritising an element of prime importance and that requires initial attention from advanced learners (i.e. lexicon), I do not neglect the fact that it is advantageous to focus attention on both syntax and vocabulary though the language acquisition journey, but rather underscore that this emphasis is intended to address the needs of the learners at such an advanced level.

2.5.1 Language competence

This thesis argues that greater focus should be directed towards the lexicon of Dickens, with this position being justified from two directions. Firstly, the grammatical rules in English are far more limited than the volume of lexicon in any given genre, even before beginning to consider the vast lexical depths of the English language in general. This suggests that when learners of English attain advanced levels, their grammatical competence in its narrow sense (syntactic knowledge and performance) and awareness of the language structure can be fairly classified as being advanced. That is, the accuracy of their output is advanced in terms of producing relatively accurate grammatically coherent English. Conversely, in the case of the lexicon competence this would appear to be limited, since it is unlikely that non-native speakers will have mastered the lexical level that renders them close to native-like levels through merely studying for three years out of four (e.g. a BA programme of four years) before they are classified as advanced learners. Furthermore, from the above analysis, the readers classified at the advanced level according to two prominent publishers do not exceed the ceiling of 3,000 headwords

in their vocabulary size, although the dispersion of such headwords may vary from one publisher to another (see Leech, Rayson & Wilson 2001; Nation 2012).

Further justification for the focus on Dickens's lexicon is the established relationship between vocabulary size and comprehension. In order to assist learners in coping with authentic texts, the extent of their vocabulary knowledge has to be increased. It has been reported that 'vocabulary size scores correlate well with and predict scores in formal writing, reading comprehension and grammatical accuracy' (Meara & Milton 2003, cited in Milton 2009: 171). Moreover, Bernhardt (2011) points out that when placing any focus on the distinctions between reading by native and non-native speakers, the depth of vocabulary knowledge and ability to comprehend highlight most clearly the differences between the former and the latter, while Milton (2009) underscores the 'strong relationship between text coverage and comprehension; that the more words you know, the better you will be able to understand when reading or listening in the foreign language' (Milton 2009: 47). Other studies also confirm this relationship between vocabulary size and comprehension, including Wagner, Muse and Tannenbaum (2007) and Qian (2002). Nagy (2007) asserts the 'correlation between vocabulary knowledge and reading comprehension' (Nagy 2007: 52), while Carlisle (2007) considers the '[b]readth and depth of word knowledge [as being] a key factor in reading comprehension' (Carlisle 2007: 78). **Lexical competence refers to the learner's ability to perceive and utilise a lexical item, inclusive of engaging with the breadth and the depth of vocabulary knowledge. The breadth of vocabulary knowledge (vocabulary size) denotes the quantity (number) of known lexical items, while the depth of vocabulary knowledge describes the extent of the learner's knowledge of the various aspects of a given word, i.e. how well these words are known (Qian 2002). These two dimensions of knowledge are measurable and considered when analysing the DCC, with the aim of assisting learners in the reading and comprehension of Dickens's work. Since these studies have all established a consistent relationship between the size of vocabulary**

and reading comprehension, this awareness can prove beneficial for instructors who teach and engage with reading in order to direct their learners' attention towards the core elements that can assist them in achieving the comprehension of given texts.

Questioning the relationship between grammaticalness and interpretability, Widdowson (1975) argues that any understanding of a literary text does not solely rely 'on the reader's knowledge of the language system or code' (Widdowson 1975: 32–3), and that it is common to encounter deviations from such systems that can be understood by readers. It is thus safe to assert that while the grammar itself can scaffold the readers' understanding of the texts, it cannot in itself be sufficient to assist (non-native) readers' interpretation of the literary work as a discourse. The distinction between *text* and *discourse* in this regard is that the discourse is a text in context, where 'linguistic elements function to communicate effect' (Widdowson 1975: 33). A more specific term that describes the functionality of *text in use* (discourse) is Lewis's (1993) *co-text*, which refers to the 'linguistic environment' (Lewis 1993: 80); that is, the 'textual context' of a given text, as described by Brown and Miller (2013: 114). Contrary to the notion of co-text is *context*, which refers to the 'situational factors' (Lewis 1993: 80), including the 'general knowledge of the world held by speakers and hearers' (Brown & Miller 2013: 107). It is evident that meaning can be understood succinctly by ensuring that the readers' grammatical knowledge is able to comprehend the basic structure of the literary texts that matches the elementary knowledge of the language code. This is the first foundation from which the reader can progress towards comprehending the discourse, providing that they are equipped with the necessary lexical knowledge. The studying of Dickens's lexicon will be conducted in Chapter 5 in regards to its co-text with the help of corpus linguistic techniques, which assist in the study of vocabulary in its co-text (see Gilquin 2010: 169).

Likewise, and as Widdowson (1975) suggests, grammar is insufficient to enable the full understanding of a discourse, despite it of course facilitating the

reader's understanding of the text. Nevertheless, if the grammar does not enable the reader to understand a discourse (i.e. a text in use), then it is unlikely that more sophisticated meanings related to a given culture can be comprehended through merely the knowledge of the code or the grammar of the language. This stance becomes increasingly evident when it is applied to non-native readers of English at advanced levels, where there is a lack of vocabulary knowledge and the cultural background necessary to respond to some of the triggers found in the discourse, unless it is instructed by another, perhaps more overt medium. This argument for the significance of lexicon for advanced learners can be correlated in the words of Lewis (1993), where he states: '[I]anguage consists of grammaticalised lexis, not lexicalised grammar' (Lewis 1993: vi), **as addressed above (see Section 1.2.2).**

2.6 Conclusion

This chapter has demonstrated why Dickens's works as authentic materials, which were not written with non-native readers or teaching purposes in mind, **result in the need for non-native readers** to deal with them differently. Identifying the source of difficulty confirms the importance of tackling this issue from another perspective, particularly when dealing with Dickens's works which are deemed to be complex and may be seen as relatively dated to learners of the English language in the contemporary arena. The justification for this necessity is the fact that authentic texts 'embody characteristics that specially-devised teaching materials often fail to capture or which they distort', as Rixon (2000: 68) suggests. In order to tackle the challenges encountered in reading Dickens's works, the focus of my study will be on lexical items and the typical usages found in his texts. Aluísio et al. (2008) affirm that the utilisation of 'low-frequency words, among other things, increase[s] text complexity' for non-native readers (Aluísio et al. 2008: 16). Regardless of the extent to which this study argues for the need to assist non-native readers in comprehending Dickens's works as authentic texts by focusing on Dickens's lexicon, it is by no means claimed that mastering vocabulary will automatically ensure full comprehension. As a matter

of fact, *comprehension* as a cognitive process can be affected by other related factors such as the reader's 'language proficiency, reading proficiency, or background knowledge' (Crossley, Yang & McNamara 2014: 96), which cannot be covered in one study addressing a complete corpus of Dickens. In Chapter 5, the investigation will focus on Dickens's works by aiming to provide an enhanced understanding of the level of complexity in lexical terms, and how non-native readers can approach such tasks with greater confidence. **By no means does this undermine other crucial factors that are associated with reading and comprehension, which alongside lexical items include the grammar and sentence length, the syntax and discourse markers, reference clarity, the progression of subordinate clauses in a linear fashion, avoiding any embedding that is complex in nature, and the equal weighting of new and given information and style components (see the Oxford Bookworms syllabus).**

Chapter 3 Stylistics, Corpus Linguistics and Corpus Stylistics

3.1 Introduction

This chapter discusses the underpinning theoretical concepts of *corpus stylistics*, which will form the framework for the analysis of the Charles Dickens Complete Corpus (DCC) to be conducted in Chapter 6. It will also detail the theoretical concepts and premises upon which the analysis of Dickens's work will be based. This will be followed by the development and discussion of the evaluation criteria necessary for the expediency of corpus stylistic analyses. The goals of corpus stylistic analysis will then be established, followed by a brief explanation of the methodological techniques employed to achieve these goals.

3.1.1 Stylistics and linguistics

Regarding the *style* embedded in language as viewed from the field of stylistics, it manifests from motivated decision making at particular levels of language, inclusive of its lexicon, semantics, syntax and other characteristics indicated, besides other 'combinations of stylistic factors' (Leech & Short 2007: 36) such as phonology, graphology, pragmatics' frequencies and collocations. Leech and Short (2007) define style as 'the way in which language is used in a given context, by a given person, for a given purpose, and so on' (Leech & Short 2007: 9), while Studer (2008) considers 'the notion of style as a motivated choice of linguistic strategies applied to induce specific effects' (Studer 2008: 7). According to Stockwell (2006), 'style can be seen as the characteristic pattern of choices associated with a writer's or projected character's "mind-style," or the pattern associated with particular periods, genres or literary movements' (Stockwell 2006: 746). Since the comprehension and interpretation of texts is guided by the determination of linguistic categories, then employing corpus

linguistics as a methodology to identify such features that trigger understanding is a valuable aim to pursue.

Mullany and Stockwell (2010) consider that style has the potential to be broadly defined as those sequences of text than can be identified due to their linguistic and discoursal characteristics. Furthermore, they consider style to encompass a range of parameters inclusive of 'every sound, word, syntactic structure, co-referential link and overall shape of the text' in existence (Mullany & Stockwell 2010: 43). The result of such choice exposes shades of meaning with contrasting emphasis, a variety of tone and evaluative influences, contrasting viewpoints and emotional traits, sincerity, perseverance and worth. Essentially, notwithstanding the historical tendency to differentiate between form (linguistic structure) and content (decoded meaning), it is not feasible in practice to separate them (Mullany & Stockwell 2010). Leech and Short (2007) **suggest** that 'the goal of [stylistic studies is] explaining the relation between language and artistic function' (Leech & Short 2007: 11). Thornborrow and Wareing (1998: 2) identify what they refer to as 'key aspects of stylistics':

- the use of **linguistics** (the study of language) to approach **literary texts**
- the discussion of texts according to **objective criteria** rather than according purely to subjective and impressionistic values
- emphasis on the **aesthetic** properties of language (for example, the manner in which rhyme can offer pleasure)

(Thornborrow & Wareing 1998: 2; bold in original)

According to Semino (2011), stylistics '[i]n its broadest sense ... is concerned with the description and interpretation of distinctive linguistic choices and patterns in texts' (Semino 2011: 541). Mullany and Stockwell's (2010) perception of the field of stylistics is one that considers the interconnections of language patterns and how they are understood. They go on to assert that despite stylisticians considering the full range of available texts, the particular focus of stylistics is on literary works,

representing the most prestigious form of language use. In respect to the development of stylistic studies, Trask (2007) asserts that

early stylistics in the 1960s and 1970s focused on semantic and syntactic aspects of poetry, largely; the 1980s saw an expansion drawing on pragmatics and discourse analysis into extended prose fiction and drama; the 1990s and recent work has drawn heavily on cognitive science to produce a cognitive poetics and on computational linguistics to produce a corpus stylistics.

(Trask 2007: 280–1)

It is valuable to consider that stylistics is no longer exclusive to literary texts, but extends beyond these to encompass all manner of texts or other text-based works such as films (e.g. see McIntyre 2010). Simpson (2004) explains why stylistics places such importance on language in interpreting texts when he states:

Stylistics is a method of textual interpretation in which primacy of place is assigned to *language*. The reason why language is so important to stylisticians is because the various forms, patterns and levels that constitute linguistic structure are an important index of the function of the text. The text's functional significance as discourse acts in turn as a gateway to its interpretation. While linguistic features do not of themselves constitute a text's 'meaning', an account of linguistic features nonetheless serves to ground a stylistic interpretation and to help explain why, for the analyst, certain types of meaning are possible.

(Simpson 2004: 2)

Simpson (2004) thus emphasises the relationship between the interpretation of the text through stylistics and the manner in which language is employed by the author, and while the linguistic characteristics may not reveal the intended meaning per se, they do offer valuable potential to consider the potential meaning(s) through comprehension of the author's prevailing linguistic features.

Bradford (1997) believes that '[s]tylistics can tell us how to name the constituent parts of a literary text and enable us to document their operations' (Bradford 1997: xii). In addressing the activity of stylistics, Bradford (1997) states:

Stylistics might thus seem to offer itself as an easily definable activity with specific functions and objectives: Stylistics enables us to identify and name the distinguishing features of literary texts, and to specify the generic and structural subdivisions of literature.

(Bradford 1997: xi)

Verdonk (2002) **believes that stylistics** embraces ‘the study of style, which can be defined as the analysis of distinctive expressions in language and the description of its purpose and effect’ (Verdonk 2002: 4). Trask (2007) indicates towards the notion that stylistics is the application of ‘the analytical techniques of theoretical linguistics to the elucidation of literary works and to the examination of the aesthetic aspects of language generally’ (Trask 2007: 280).

Cook (2008) advocates positioning stylistics within the domain of text analysis studies **through an exploration of Stubbs’s (2005) beliefs** by addressing issues related to stylistics as being *sciences* which ‘[seek] to be descriptive rather than prescriptive, replicable by other analysts, expounding objective facts about language use’ or representing *art* that ‘evaluates and prescribes, imposing the writer’s views upon the external world, saying as much about the analyst as the analysed’ (Cook 2008: 305). Other issues related to the positioning of stylistics explored by Cook (2008) include the distinction between *literary criticism and stylistics*, and also the difference between *stylistics* and *quantitative stylistics*, and the issue of *semantic/discourse prosodies’ evaluation or calculation*. The concept of style utilised by corpus stylistics, and the link between stylistics in general and corpus stylistics in particular will be explained later in this chapter, in section 3.4.

In its general sense, stylistics is ‘the study of style’ (Leech 2008: 54; Wales 2011: 399). Leech and Short (2007) define the nature of the study by considering stylistics as ‘the (linguistic) study of style’ (Leech & Short 2007: 11). The two key stylistics-grounded concepts arising from these definitions are *linguistic study* and *style*.

In this context *linguistic study* is concerned, as Busse and McIntyre (2010) observe, with 'the intra-linguistic features of a text' (Busse & McIntyre 2010: 6). Moreover, as Jeffries and McIntyre (2010) clearly state, the study of stylistics 'uses models of language, analytical techniques and methodologies from linguistics to facilitate the study of style in its widest sense' (Jeffries & McIntyre 2010: 1). In contrast to literary studies, where style refers to a broader range of features so as to signpost the 'distinctive patterning of language associated with an **author**, movement or period amounting to a "verbal fingerprint" or "verbal trademark"' (Montgomery et al. 2007: 359; bold in original), the style in stylistic studies primarily concerns itself with the linguistic features of the text. As Wales (2011) asserts, this focus on the linguistic features of a text implies that

'[t]he goal of most stylistic studies is to show how a text "works": but not simply to describe the FORMAL features of the texts for their own sake, but in order to show their FUNCTIONAL significance for the INTERPRETATION of the text; or in order to relate literary effect or themes to linguistic "triggers" where these are felt to be relevant.'

(Wales 2011: 400; emphasis in original)

Jeffries and McIntyre (2010) reformulate this view by stating that 'analysing style means looking systematically at the formal features of a text and determining their functional significance for the interpretation of the text in question' (Jeffries & McIntyre 2010: 1).

3.1.2 Possible challenges for non-native readers

One of the challenges that non-native readers may face when engaging with literary texts such as Dickens's works is the potential of having 'variant readings' for the same texts (see Carter 2010: 64). In the context of Dickens's works for instance, as socially literary texts they may be characterised by portraying a number of aesthetic and literary features aimed at creating specific responses, feelings or to draw certain

images to the reader's mind and conscience. Engaging with such texts is not subject to specific measures by which one particular meaning can be elicited, but rather that other intertextual and intratextual factors can play a role in the process of reading and creating various meanings. The linguistic structures found within the literary texts are certainly not without function, but rather they are employed to lead to multiple responses ranging from enhancing the text's enjoyment to conveying a specific message. To offer an anecdotal example from teaching a course in Translation in order to shed light on how non-native readers may be distracted by a lexically possible meaning, a statement such as 'the John F Kennedy School of Government at Harvard' was misunderstood as **being** a public/governmental institution for educating children which had been named after John F Kennedy. Louw (2007) suggests that the 'literary world of any text is assembled afresh every time that text is read' (Louw 2007: 104), while Birch (1989) asserts:

There are other classifications, other readings, other ways of articulating how ... text means ... There is no single text with a single meaning. Meaning is relative to ideology, and the way we classify a text as 'working' in a particular way says a great deal about the ideologies we are practising—consciously or otherwise.

(Birch 1989: 28)

The response to why it is possible to obtain different readings for the same literary text lies in recognising the contributing factors in creating meaning. In the case of Dickens's works, a combination of circumstances (personal, historical, political and social) may dictate how meanings are expressed by the author. Likewise, the reader's personal circumstances also influence the manner in which the meanings are constructed from the text and cognitively conceived, as retrieval of meaning differs from one reader to the next. Comprehending all the possible readings of Dickens's works is not a realistic goal, and therefore the purpose of this study does not focus on the exploration of how variant readings might be accounted for in Dickens's works, as this objective falls beyond the scope of this study. It is noteworthy that stylistics assumes there not to be an infinite number of

interpretations of a text, but rather a finite number of reasonable interpretations that are based on evidence primarily from the text itself. Thus the requirement to confine this study to what can be referred to as the 'basic meanings' or 'primary meanings' is confirmed, particularly if the lexicography term can be quoted with regard to reading Dickens's work.

Comprehending the basic meanings of Dickens's texts will therefore form the platform from which other readings can start. This claim that a text can result in different readings is plausible if the three factors relating to the reading process are taken into consideration. These factors are as follows: firstly, the text as a source of meanings, including its context (pragmatic factors); secondly, the manner of reading (i.e. how meanings are accessed and retrieved, for example surface or deep reading, comparison and determining the characters of the texts – similarities, truthfulness, falsehood, creativity in a fiction, norms (the conventional way of expressing an idea or describing a situation), foregrounding, contradiction and ambiguity); thirdly, the reader/receiver of that meaning, and his/her culture, awareness, previous knowledge, schemata and imagination, which all combine to influence the absorption of the meanings of a text. On the other hand, it is commonly accepted in literary pedagogy that literary works can have multiple readings that aim 'not to determine which of the multiple readings was "right," per se, but rather to investigate the different literate practices that produced these multiple readings' (Miller 2003: 132). Beard (2003), for instance, suggests that rereading a literary work promotes a 'sense of opposition to a dominant reading position [which] is one of the main ways in which multiple readings of texts are possible, often through exploring the context of a text's production and reception' (Beard 2003: 43). Jaszczolt (2005) provides a linguistically 'adequate formal account' of constructing the meaning using the framework of 'Gricean and post-Gricean pragmatics' (Jaszczolt 2005: xi).

3.1.3 The historical dimension in Dickens's works

Another challenge that readers may face relates to the fact that the texts of Dickens based on his last written work, *The Mystery of Edwin Drood*, which he died before completing in 1870, were written over one hundred and forty years ago. Therefore, the environment and general contexts of Dickens's works represent a different world from that of the modern era. In discussing Dickens's London, Hawes (2007) states that 'Dickens wrote in a uniquely vivacious fashion about its scenes and people' in his different works (Hawes 2007: 25); moreover, 'London appears somewhere in all his novels, except *Hard Times*' (Hawes 2007: 25). Among the relevant challenges is the historical dimension of the text. Albeit in relation to the 'Elizabethan world', Culpeper (2009) emphasises the importance of considering the 'milieu' of the text, wherein the same words may have conveyed different meaning in their historical setting than they would in present-day English (historical linguistics). He proceeds to offer an example in the word 'love', which had the sense of 'friends' or 'intimates', but no strong implication of relation as found in the present-day perception (Culpeper 2009: 47). The role that corpus stylistics can play here is to identify the wider context of where the lexical item was used by way of concordance, which is then exploited to draw a more precise picture of the social setting in which the lexical item was used. By identifying similar contexts and usages of the same lexical item, this awareness will aid non-native readers who are not sufficiently cognisant of the historical culture to perceive the meaning of the texts without confusing the historical meaning with that perceived from the same lexical item when using their present-day knowledge of the word. The issue here entails how non-native readers who are unaware of the historical culture could be primed for understanding of the text in basic terms, its relation to its historical denotation and what should be understood in this regard.

3.1.4 The cultural dimension in Dickens's works

The culture manifested in Dickens's work is a further challenge to readers from other and diverse cultures and backgrounds, as Dickens was 'living in a different cultural world' (Hawes 2007: 4). Hall (2005) considers that '[l]iterature, language and culture have always been seen as interrelated in various ways', and that '[l]iterature was first conceived and taught as offering a privileged and prestigious access to distinct national "cultures" and languages' (Hall 2005: 66). Miller (2002) considers literature to be 'a feature of any human culture at any time and place' (Miller 2002: 1). Brussino and Gunn (2008) assert that '[i]solation from the target culture limits learners' exposure to authentic language and cultural input (Brussino & Gunn 2008: 1–2), and thus risks rendering the classroom a sterile or artificial study environment. Considerable effort is devoted to avoiding such potentially negative influence on learner interest, motivation and performance. In the case of Dickens's work, although the '[a]dvances in multimedia, information and communication technologies have made valuable additions to the teacher's toolkit' (Brussino & Gunn 2008: 1–2), they still can be limited when applied to Dickens's works which are chronologically remote from the cultures of the learners, as in the case of, for instance, Arab learners or readers. According to Abrams (1999), one of the '[r]elevant external references' that assist the reader in establishing 'a determinate interpretation' of literary texts is 'the author's cultural milieu' (Abrams 1999: 128–9). Meaning construction is influenced by usage. This entails situated acts of language use and other non-verbal cues, such as gestures, in service of the expression of situated, goal-directed communication intentions, in a particular physical setting and cultural milieu, making use of various cognitive mechanisms and processes. This study will not be concerned with addressing the question of how to **engage with** the cultural concepts of Dickens's historical period, although some studies have produced an affirmative answer that non-native readers of English, for instance, can be prepared to perceive a given text by providing them with the necessary **cultural**

knowledge in advance of their reading of the texts, thereby adopting the concept from schemata theory (e.g. see Carrell 1984: 334).

3.1.5 Utilising corpus stylistics to facilitate stylistic analysis

According to Lazar (1993), stylistics within the classroom 'has two main objectives: firstly, to enable students to make meaningful interpretations of the text itself; secondly, to expand students' knowledge and awareness of the language in general' (Lazar 1993: 31). Stylistics has therefore the potential to 'help students to read and study literature more competently', as well as '[providing] them with excellent language practice' (Lazar 1993: 31).

The first practice to attempt when analysing text can be in extracting, for instance, the keywords of the analysed text, although Mahlberg and McIntyre (2011) caution that 'generating keywords does not constitute an analysis' (Mahlberg & McIntyre 2011: 206). Therefore, the extracting of keywords from the DCC is no more than a descriptive analysis at best; this cannot be viewed as stylistic analysis. It is for the stylistician to advance this further for stylistic framing where keywords can provide stylistic insight to support textual interpretation. This can be illustrated by the manner in which stylisticians typically treat the phenomenon of keywords in the text. Extracting the keywords from the node corpus as an initial step represents a corpus linguistic exercise. Nevertheless, stylisticians often proceed beyond this in a quest to discover the meanings that can be elicited from such lists of keywords. The manner in which keywords can be employed to determine with objectivity those words that feature in a corpus with frequency, and to then allow comparisons to be made with larger reference corpora has been stated by Hunt and Carter (2012). Moreover, it is possible to extend the analysis of the keywords list still further by analysing specific grammatical forms in order to better understand their functions in the texts; for example, how adjectives contributed towards establishing the settings

of, or scenes in, a literary text, as in one of Dickens's novels. Further examples of these interpretations will be demonstrated in Chapter 6.

As for the value of corpus stylistics, in his discussion of James Joyce's short story, *Eveline*, O'Halloran (2007) highlights that 'through use of corpus techniques of analysis ... arbitrariness can be reduced, since [the] interpretations have been based on objectively revealed features of the text' (O'Halloran 2007: 241; see also Stubbs 2005 and Hunt & Carter 2012). O'Halloran (2007) further considers that 'interpretations [can be] constrained in a principled way, since they have been based around keywords revealed through corpus-informed methods ... [though] there are, of course, other possible interpretations of the keywords' (O'Halloran 2007: 242). The employment of keywords will be discussed in greater detail in Chapter 5.

Another advantage of the method of corpus stylistics as suggested by Stubbs (2005) is that its 'findings ... sometimes document more systematically what literary critics already know (and therefore add to methods of close reading), but they can also reveal otherwise invisible features of long texts' (Stubbs 2005: 22). The tendency to 'investigate the validity of the claim[s]' found in critical literature is not a recent phenomenon, as Davidson (1990) reports on the use of corpus linguistics in teaching the English language, whereby corpus linguistics is intended 'to help [students] investigate usage in written and spoken English, vocabulary statistics, and literary-critical claims based on statements about the language of texts' (Davidson 1990: 84).

Hunt and Carter (2012) point out that '[c]orpus stylistics provides new insights into narrative texture and demonstrates the importance of recurrent linguistic features in shaping meaning' (Hunt & Carter 2012: 27). Consideration of a further practical feature of using corpus stylistics is that it can manage a large collection of texts, as Hunt and Carter (2012) suggest, with the aim to 'typically explore interpretations of literary texts identified in conventional readings but which may be hard to substantiate by traditional methods' (Hunt & Carter 2012: 30).

Mahlberg and McIntyre (2011) highlight the benefit of employing corpus stylistics in stylistic analysis as it 'enables us to address what has long been an issue with the analysis of prose fiction ... the problem of length and the fact that most prose texts are simply too long for the stylistician to deal with' (Mahlberg & McIntyre 2011: 205; see also Leech & Short 2007: 2). Mahlberg and McIntyre's (2011) article 'contribute[s] to this emerging field by demonstrating the interplay of qualitative and quantitative methods in a corpus stylistic analysis of Ian Fleming's novel, *Casino Royale*' (Mahlberg & McIntyre 2011: 204).

3.2 Stylistics, *its aim and objectivity*

In his brief history of stylistics, Stockwell (2006) considers how 'stylistics can be seen as a direct descendant of rhetoric', subsequently discussing the 'three direct influences which produced stylistics', as suggested by Fowler (1981), namely: 'Anglo-American literary criticism; the emerging field of linguistics; and European, especially French, structuralism' (Stockwell 2006: 743).

The aim of stylistic investigation is to distinguish and understand the linguistics-related features of the text that direct towards the essence of the (literary) impact. Therefore, stylisticians primarily restrict their concerns to considering the text itself in isolation from all that may exceed the text or extend beyond its inherited features. Rather, the stylistic approach aims to distinguish those **linguistic characteristics that distinguish literary discourse and how its effect and meaning are created. Unsurprisingly, stylistics aims to develop methods and approaches that enable other analysts to recognise precisely those procedures by which the patterns and features of a text can be identified and interpreted, thus raising awareness of how such linguistic characteristics function in a given text and then construct a message (i.e. create meaning). Despite stylistics being concerned with the dimensions of meaning and effect, the focus during the analytical process is primarily on the axis of *selection and combination*, as termed by Jakobson (1960), when**

he states that 'the poetic function projects the principle of equivalence from the axis of *selection* into the axis of *combination*' (Jakobson 1960: 358, cited in Simpson 2004: 52; emphasis added). Selection refers to choosing from equivalent options of lexical items, while combination is concerned with constructing the syntactic links, that is, the grammatical organisation.

Stylistics is an analytical method that avoids impressionistic judgments and rather follows the approach of scientific description obtained from the field of linguistics. It is sufficient to be clear that stylistics intends to establish a logical approach to that which is primarily identified by intuition. That is, the first election of a given text will essentially be purely subjective and intuitive, but the stylistician intends to progress from that stand point to locate objective features to justify the merits of that text. Short et al. (1998) state that:

[Stylisticians] know they can be wrong, and make a point of making clear the grounds for their views, and how they could be falsified, precisely so that others can challenge and test them, thus advancing our knowledge about texts and how we understand them. For a stylistician, then, being *objective means to be detailed, systematic and explicit in analysis*, to lay one's interpretative cards, as it were, clearly upon the table.

(Short et al. 1998: 46; emphasis added)

In explaining Wales's (2011) sense of stylistics' analytic objectivity, Carter (2010) deems objectivity as 'being methodical, systematic, empirical, analytical, coherent, accessible, retrievable and consensual' (Carter 2010: 61). If these characteristics are to be obtained through corpus stylistics, then an analysis must also be provided that is accurate, precise, duplicable, materialised and (most importantly) text-based-evidenced proof. In other words, it can be considered to be irrefutable. Thus, if a new interpretation or reading is provided to a given text, then it will be equipped with concrete evidence from the text from which it was computationally elicited (Carter 2010: 61).

Kraus (2008), in reference to researchers from structuralist schools, observes that such researchers have deemed stylistics to be an independent field dedicated to consideration of the selection and placement of units of language, which can then be comprehended objectively in order to identify a hierarchical structure of communication in both written and spoken forms. The aforementioned units can point towards the range of levels present in language and speech, such as 'the lexical, syntactic, metric, phonetic' (Kraus 2008: 1012).

Kraus (2008) further asserts that in the view of Jakobson (1960), the role of stylistics is solely objective, *since it examines the characteristics of internal structures of the text*. This implies that *stylistics is concerned with the relations between the text's internal structures on different tiers* such as the lexical, syntactic, metric and phonetic levels. Jakobson (1960) describes that the poetic purpose of language is achieved through understanding the precise relationship that exists between the *selection* and *combination* operation, whereby *selection entails the process of choosing the language means from a particular paradigm of equal means*, while *combination* indicates how these selected language means are placed on a syntagmatic axis. Meara (2009) defines syntagmatic associations as 'associations that complete a phrase (syntagm)', offering the examples of 'BRUSH teeth, HOLD hands, BLACK mark and BANK robber' (Meara 2009: 6, capital in original).

One potential rationale for the misconceiving of the concept of objectivity in stylistics is that the opponents of this concept consciously or unconsciously define the *style* through a narrow meaning. That is, they define style as per literary studies and then judge the concept of objectivity by exporting their own notion of the term *style* into stylistics. *In literary studies, style is defined as the 'distinctive patterning of language associated with an author, movement or period amounting to a "verbal fingerprint" or "verbal trademark"'* (Montgomery et al. 2007: 359). This can be clarified by first raising awareness of the concept of style within stylistic studies, as it has now been developed to encompass without reservation all the linguistic features

of the text (see Mackay 1996; Short et al. 1998; Mackay 1999; Short & van Peer 1999). In literary studies, the reference to *style* is not directed to the linguistic features by the open and free choice the producer of the text has made.

Nevertheless, to maintain objectivity, stylistic analysis relies on identifying the linguistic features of the text, while the intuitions in the corpus stylistic approach are removed from the process of determining the data that attract the attention for study. This is explained by the analysis being guided through the frequency information of the text. Fischer-Starcke (2009) defines '[c]orpus stylistics, as a combination of corpus linguistics and stylistics which uses Jakobson's (1960) and Halliday (1971) premises in analysing the syntagmatic axis of a text' (Fischer-Starcke 2009: 494). When stylistically examining texts, the emphasis on objectivity relies on the employment of linguistic knowledge, addressing the relationship between the forms and meanings, and removing the individual responses to texts as a means of their analysis. This is not to fully deny the role of readers' personal responses, but it is barely possible to prove and trace scientifically, where possible, the various responses in the analysis of literary or non-literary texts.

Simpson (1993) states that stylistics 'practitioners use linguistic analysis as a basis for their interpretations of texts' (Simpson 1993: 2). The claim of relative objectivity in stylistic studies comes from the 'reliance on the "science" of linguistics', as Simpson (1993: 3) describes it. The exponents of relative objectivity base their interpretation of the texts on the linguistic features of the texts, which can by no means be determined subjectively as these features stand independently in the texts. The opponents of this claim may still confront it by asserting that it is both the selective choice of these linguistic features and the subjective interpretation that we see as subjective interpretation of the texts, as conducted in literary studies and criticism. Stylisticians here again objectively respond by stating the fact that no stylistician claims *pure objectivity* in the textual analysis, rather '[t]hey prefer to recognize instead that all interpretations are in some sense context-bound and are

contingent on the position of the analyst relative to the text', as Simpson (1993: 3) clearly states. Rayson (2008) believes that the actual texts inform the grounds with which to interpret an author's style, which stands independently of interpretations based merely on the impressions of the analyst:

Another direction in the notion of objectivity is that the stylistic interpretation of texts is supported and based on evidence from the texts themselves, which is discrete from providing explanations with no intertextual evidence or grounded solely on intuition or feelings.

(Rayson 2008: 520)

Further indication to the objectivity of stylisticians is asserted by Simpson's (2003) claim that objectivity 'is not to deny the usefulness of subjective evaluation in critical interpretation, nor is it to advance a claim for total "objectivity" in critical practice' (Simpson's 2003: 57).

A part of, or one way of expressing the objectivity in stylistic studies is that stylisticians are objective in claiming the notion of objectivity in stylistics, and that rather than asserting any absolute objectivity, they recognise and make clear any subjective decisions that are taken during the process of studying and analysing texts. Therefore, stylisticians assert no objectivity in the initial step, which involves selecting the text for analysis, but rather, Crystal and Davy (1969) '[emphasise] right away that the first step in any stylistic analysis must be an intuitive one' (Crystal & Davy 1969: 12). Another stage where stylisticians claim no objectivity is as Semino (2011) highlights in the 'interpretation and analysis' stage, whereby the analysis 'inevitably relies on the analyst's intuitions and knowledge of the data' (Semino 2011: 543). Crystal and Davy (1969) consider objectivity to be 'the stylistician's main competence' as they

interrelate his observations within the framework of some theory, and thus piece together any general pattern of linguistic variation which may exist. This is where the objectivity claimed for stylistics comes in, not in the initial step, which is [a] wholly subjective (though, one trusts, informed) decision.

(Crystal & Davy 1969: 12)

Semino (2011) finds that objectivity in stylistic studies is linguistically oriented, as 'it is based on the application of relevant theoretical frameworks and on a systematic understanding of patterns of variation in language use' (Semino 2011: 543). It is further asserted by Semino (2011) that 'the confident tone of some work in stylistics may be mistaken for the belief that linguistic analysis enables one to arrive at the "true" or "best" interpretation of a text. Most stylisticians seem to agree that they do not aim for some general (and indeed suspect) notion of objectivity, but rather that their goal is to produce textual analyses that are explicit, rigorous, systematic and replicable' (Semino 2011: 549).

Upon establishing the notion of objectivity in stylistic studies, I view it as *stylistics' relative objectivity, asserting that stylistics promotes objectivity rather than subjectivity, and that it is relative, not absolute*. Objectivity is considered to be one of the most important pillars of substantive academic research, where academic standards are maintained in all stages (if possible, commencing from the identification of the research questions and goals, and the means and methods that lead to the desired results, in order to arrive at objective outcomes based on an independent analysis). This is anticipated to prevent the research from stating personal impressions, though subjectivity cannot be completely eliminated in determining the nature of the search questions and reported results. This relative objectivity is the valuable qualification in the stylistic analysis of texts.

Mackay (1996) and Short et al. (1998), and then Mackay (1999) and Short and van Peer (1999) debate the concept of objectivity in stylistics' studies. Referring to and reporting on the perspectives of Leech and Short (1981, 2007), in disagreement Mackay (1996) claims that 'complete objectivity is unattainable, a less than complete objectivity is better than none at all' (Mackay 1996: 82). Moreover, Jeffries (2000) suggests 'that no science is completely without premises (fundamental assumptions) and no scientific data is amenable to absolute proof', seeing that '[r]elative objectivity is desirable in stylistic and related endeavours' (Jeffries 2000: 11). Leech

and Short (2007) assert that 'without quantitative confirmation, statements on style lack the support of concrete evidence' (Leech & Short 2007: 57), while Short et al. (1998) assert that 'we cannot expunge our personal response from our analyses, and would never want to' (Short et al. 1998: 46). For stylistic analysis, 'in basing its investigations on observable data, it offers a greater level of objectivity than other methods', as Huckin (2004: 29) observes.

It appears that one of the triggers of the criticism directed towards the notion of objectivity in stylistic studies is 'the lack of an agreed methodology (or methodologies) which can be replicated by different researchers', as posited by Jeffries (2000: 11). Furthermore, she can find 'no reason why we should not work towards both methodological refinements and improved objectivity, whatever the approach, model or method that we adopt' (Jeffries 2000: 11). As Short and van Peer (1999) explain:

Stylisticians can try to be precise about what they are claiming, to make their analyses as detailed and precise as they can, to be systematic in what they do, not allowing themselves to sweep counter-evidence under the carpet, to avoid making un-understandable and unfalsifiable statements, to be as inclusive as they can when undertaking the stylistic analysis of a text by making sure that they look at all the aspects of language and pragmatic processing that are available to them, confront their hypotheses with the real responses of real readers etc. This is what we mean by trying to be objective.

(Short & van Peer 1999: 273)

Stylistic research is concerned with the linguistic dimension of texts in a scientific attempt to address the essence of the literary impact that can be acquired solely by linguistically identifying formulations of the text. Therefore, stylistics limits its focus to the text itself, notwithstanding any attempt to exceed that text by seeking, for instance, the historical or psychological circumstances.

3.3 *Corpus linguistics*

Regarding the question as to what corpus linguistics is, Leech's (1992) definition remains relevant to many corpus linguistic studies when he states **that**

“corpus linguistics” refers ... to a methodological basis for pursuing linguistic research. In principle (and often in practice) corpus linguistics combines easily with other branches of linguistics: we can study phonetic[s], syntax, sociolinguistics, and any other aspect of linguistics by means of corpora, and when we are doing this we can be said to be combining techniques of corpus linguistics with the subject-matter of phonetics, syntax, sociolinguistics and so on.

(Leech 1992: 105–6)

Addressing a further aspect of corpus linguistics, McEnery and Hardie (2012) ‘define corpus linguistics as dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions’ (McEnery & Hardie 2012: 1). Furthermore, Bennett (2010) describes that it ‘approaches the study of language in use through corpora ... A corpus is a large, principled collection of naturally occurring examples of language stored electronically’, continuing that ‘corpus linguistics serves to answer two fundamental research questions’ (Bennett 2010: 2):

1. What particular patterns are associated with lexical or grammatical features?
2. How do these patterns differ within varieties and registers?

(Bennett 2010: 2)

These deflections represent a perspective of corpus linguistics as a methodology (see Meyer 2002; McEnery & Hardie 2012) in contrast to being a theory in its own right, as discussed by Gilquin (2010: Chap. 2), who questions whether corpus linguistics is a theory or a methodology. Leech (1992) argues ‘that computer corpus linguistics ... defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject’ (Leech 1992: 106). Moreover, while Taylor (2008) considers the question of “‘What is corpus linguistics?’ – is it a discipline, a methodology, a paradigm or

none or all of these?', she 'does not attempt to offer any definitive answers' (Taylor 2008: 179). Biber (2011) concludes that the 'research carried out on a corpus has the goal of describing the patterns of language use in the target textual domain' (Biber 2011: 15). In fact, he extends this position by arguing for the corpus as offering the most appropriate approach to considering textual fields, with corpus analysis being representative of the most effective and rigorous empirical means by which to analyse the application of language employed in any particular domain (Biber 2011). Biber et al. (1998: 4) observe that 'corpus analysis generally shares four characteristics':

1. it is empirical, analyzing the actual patterns of language use in natural texts;
2. it utilizes a large and principled collection of natural texts, known as a "corpus," as the basis for analysis;
3. it makes extensive use of computers for analysis, using both automatic and interactive techniques;
4. it depends on both quantitative and qualitative analytical techniques.

(Biber et al. 1998: 4)

Further details on corpus linguistics can be found through (McEnery & Hardie 2012; McEnery & Wilson 2001; McEnery & Gabrielatos 2006; Louw 1997; Baker, Hardie & McEnery 2006; Hunston 2006; Hunston 2011: 4–5). Finally, Toolan (2006) describes that '[i]n the effort to bring corpus linguistic tools to bear on literary linguistic analysis, many recent publications have begun to map out new possibilities: see, among others, Stubbs (2001, 2005); Sinclair (2004); Hori (2004); Scott (2006); Scott and Tribble (2006); Short and Semino (2003); Hoover (1999); and Hoover et al. (2006)' (Toolan 2006: 181).

Biber et al. (1998) **mention** the importance of underscoring that the analyses of corpus must extend beyond mere calculations of linguistic characteristics, to ensure that the qualitative functional understanding of the quantitative data is achieved. They stress that rather than reporting quantitative data, the aim of corpus-based study should be to investigate the relevance of these findings to promoting enhanced

understanding of language use patterns. Moreover, they state that complex methodological issues can arise when attempting to establish such patterns and examining contextual elements, as it is common for humans to recognise the atypical rather than the typical, and thus results grounded in intuition can be flawed. Biber (1998: 6) refers to lexical associations, namely to investigate how a lexical item is behaving and how it is systematically associated with other particular words. Taylor (2008) explores the discussion regarding the definition of corpus linguistics, inclusive of what discipline it falls under, with his preliminary investigation revealing that 'there is a multiplicity of views available from some of the most influential corpus theorists' (Taylor 2008: 182). He states:

In terms of what corpus linguistics 'is', not only have various definitions been offered, but alternatives have been explicitly addressed and rejected. These include, as we shall see: *corpus linguistics is a tool, a method, a methodology, a methodological approach, a discipline, a theory, a theoretical approach, a paradigm (theoretical or methodological), or a combination of these.*

(Taylor 2008: 180; italics in original)

Taylor (2008) reports Leach's (1992) early 1990s argument that computer-based corpus linguistics describes not only an innovative methodology for the study of language, but more significantly a contemporary and emergent philosophical approach, before proceeding to label the nature of computer-based corpus linguistics as a new paradigm. Moreover, Stubbs (1993, cited in Taylor 2008: 180) refutes the constraining paradigmatical description of corpus linguistics, asserting that in the subject's aspirations, a corpus not only represents a linguistic analytical tool, but also a significant theoretical concept in linguistics, while Teubert (2005, cited in Taylor 2008: 180) underscores the theoretical formulation, defining corpus linguistics as the study of language through a theoretical approach.

Biber (2011) asserts that corpus linguistics represents a research approach which enables the use of language to be described empirically, whereby such

research involves the analysis of a corpus, which is a significant and methodically organised digital collection of works. A corpus is a sample intended to reflect a text-based arena of a language, such as daily conversations, editorials found in newspapers, electronic personal messages or the works of Charles Dickens. As with any sample, a corpus can be explored to understand its ability to reflect the wider population, which in the literary case would be the textual arena. Therefore, research conducted on a corpus has the objective of identifying language use patterns in specific textual arenas.

Gries (2009a) points out that 'the notion of "corpus" refers to a machine-readable collection of (spoken or written) texts that were produced in a natural communicative setting, and the collection of texts is compiled with the intention to be representative and balanced with respect to a particular linguistic variety or register or genre and to be analyzed linguistically'.

While addressing the concept of corpus linguistics, McEnery and Gabrielatos (2006) indicate how some 'assert that corpus linguistics is not a branch of linguistics, nor a linguistic theory, but a methodology' (McEnery & Gabrielatos 2006: 34), although others argue that 'corpus linguistics is more than just a methodology'.

Regarding the function of theory in corpus linguistic research, **McEnery and Gabrielatos (2006)** believe in the value of considering differing approaches being located within a continuum, as opposed to at one or the other end-point extremes. They posit that at one end of the spectrum the corpus is employed to reveal evidence to support or refute a considered theory or theoretical framework, while at the other the patterns of data revealed through the corpus can be utilised as a foundation from which to reveal insights into language without the confines of established theories or frameworks, with the objective of creating purely evidence-based theory. In the context of this study, the two approaches are merged, i.e. corpus-based and corpus-driven approaches to analysing Dickens's work. In Chapter 5, where several corpus

stylistic applications are outlined, the DCC is approached with no preconceived theory or intuitions regarding the work of Dickens, **but** rather it is concerned with gaining evidence-based insights into Dickens's use of language. McEnery and Gabrielatos (2006) state that 'the corpus-based approach ... is associated with corpus research influenced by the work of Leech (1991a) or Halliday (e.g. 1991) ... whereas the corpus-driven approach ... is associated with corpus research influenced by the work of Sinclair (e.g. 1991) and Firth's contextual theory of meaning' (McEnery & Gabrielatos 2006: 36), although these two approaches are not two opposing positions but rather a continuum of the corpus linguistic approaches in **textual** analysis, as thoroughly debated by McEnery and Gabrielatos (2006: 35–40).

It has been claimed that in producing text-based data, **it** is unfeasible to significantly analyse corpora without the support of a competent and intuitive analyst to gain understanding of the language (by employing native or near native speaking skills) and knowledge of the language (by utilising linguistic ability) (Leech 1991), with the assertion essentially that the use of corpora is reliant on the corpus and intuition, as opposed to the corpus or intuition. Leech (1991b) also points out that '[s]imilarly, the limitation of corpus size means that samples provided in the corpus may be statistically inadequate to permit generalization to other samples of the same kind' (Leech 1991: 100), while he 'also indicated to the size, limitation of language variety i.e. one geographical provenance and limitation in historical period' (Leech 1991: 99).

3.4 Corpus stylistics

Through merging the two approaches, the aspiration is to identify patterns and regularities regarding Dickens's language to meet one of **stylistics' goals**, which is related to comprehending literary texts. This means that corpus linguistics provides, primarily, the tools necessary to identify the basic items and forms of lexis, besides other significant patterns such as collocations and phrases, which enable the reader

to comprehend the literary texts. Thus, corpus stylistics as a term covers and combines two disciplines: *corpus* refers to the use of corpus linguistics approaches, while *stylistics* indicates the linguistic analysis of literary texts. Therefore, corpus stylistics points to employing corpus linguistics' approaches and tools utilised to realise stylistics' goals, namely to linguistically analyse such texts. Corpus methodologies are applied in large text bodies to elicit linguistic patterns and phenomena that lead to generalisations about a language in terms of its syntax and semantic behaviour (the behaviours of lexicons within a given text, e.g. Mahlberg 2009).

Corpus stylistics is an approach that employs the service or the evidence elicited from corpus linguistics to support literary analysis, transforming it from a basic, subjective 'claim-and-quote strategy' (McIntyre 2010: 180) into evidence-based objectivity that encompasses linguistic features. Carter (2010) asserts that analysis through corpus stylistics is a process that can be considered to be comparatively objective and methodological, and one which is steered through a process of interpretation, in itself comparatively subjective; moreover, although the great potential of corpus stylistics has yet to be fully realised due to the necessity of surmounting the philosophical and practical challenges, literary stylistics represents a significant methodological component of the stylistics field moving forward.

The reliance on the corpus linguistic approach in the empirical analysis of naturally occurring literary or non-literary texts becomes a commonly employed methodology in several areas of linguistic research (see Balossi 2014: Chap. 4 for a review on corpus approaches applied in stylistic studies). McEnery and Gabrielatos (2006) underscore that studies employing corpus linguistics may have a theoretical perspective, with some targeting a contribution towards a 'specific theoretical framework' in either a direct or indirect manner (McEnery & Gabrielatos 2006: 46). Such studies may aim to be positioned 'within the paradigm of lexico-grammar, probabilistic grammar, cognitive linguistics, lexicographical, and language teaching

applications', with their lens also placed on the (i) definition and exploration of lexical meaning, (ii) focus on 'the phraseology of a word', (iii) exploration of how multi-word lexical items behave, (iv) investigation of lexico-grammar, (v) consideration of the grammatical structure's lexical qualities, or (vi) analysis of the grammatical category distribution (McEnery & Gabrielatos 2006: 46). McEnery and Gabrielatos assert that methodologies based on corpora are increasingly being employed in studies within the fields of 'pragmatics and discourse analysis ..., critical discourse analysis ..., metaphor ... and stylistics' (McEnery & Gabrielatos 2006: 46).

According to Mullany and Stockwell (2010), the first decade of the 2000s witnessed additional innovations in stylistics, which were adopted with great interest. In the case of *corpus stylistics*, there has been a revolution in regards to the breadth and detail of the stylistician's potential, resulting from the increasing employment of sizeable language corpora and computer software to examine and explore them. Consequently, patterns can be identified empirically within vast bodies of text, or even the aggregated works of authors or literary movements, both consistently and effectively through automated processes. Moreover, software programmes have the potential to swiftly highlight all instances of words or phrases, as well as the contextual locations in which they are found. As a result, intuitions or assertions of texts that are beyond the realms of traditional stylistics' approaches can be investigated and examined with relative ease.

According to Leech (2007a), the term *corpus stylistics* has 'come into use since 1981' (Leech 2007: §1). Hockey (1994) draws attention to the use of concordances 'as a basis for stylistic analyses and even for studies of disputed authorship' (Hockey 1994: 682) that can be traced back to 1964, while offering an example of the study of the *Federalist Papers* by Mosteller and Wallace (1964). Mahlberg (2007a) also states that 'Hockey (2000: 67 ff.) presents examples of studies that go back to the 1960s'.

Biber and Conrad (2009: 73–4) underscore the advantages of the corpus-based approach in the analysis of texts, **while** Biber (2011) also finds that ‘the long tradition of computational and statistical research on authorship attribution and literary style’ has contributed to the development of the ‘recent innovation’ (Biber 2011: 20) of corpus stylistics, regardless of having different objectives; however, it is only recently that these types of studies have been specifically associated with corpus linguistics, conducted primarily within the scope of corpus stylistics.

The combination of the two disciplines of corpus linguistics and stylistics encompasses a methodological approach adopted from corpus linguistics that realises the achievement of the goals set by stylistics. The adaptation of the methodological approaches grounded in corpus linguistics is employed in stylistics in order to linguistically analyse literary or non-literary texts, for the reason that ‘literary texts use the same resources of language as other texts, and all those features which in the past have been associated with “literariness” can be found in other kinds of text as well’ (Haynes 1995: 58). In this regard, corpus stylistics as a term can be examined at two different levels of enquiry. The first is that *corpus* in essence represents *methodology* (Gries 2009b), while *stylistics* also essentially constitutes *concept, goals* and *linguistic enquiries*. The second level is that the *practical aspect* is sourced from corpus linguistics, while the *theories* are of stylistics. In response to a number of the original questions posed by stylisticians, such as ‘the extent to which foregrounding is quantifiable and whether authorial style really is ... distinguishable’ (Busse & McIntyre 2010: 10), McIntyre (2010: 10) suggests that these can be responded to through utilising corpus stylistic methods, which rely on advanced computer technology. The advancement of technology in being able to identify the linguistic features of a text, and for the keenness of stylistics in analysing the effect of these linguistic features of (literary) texts in general, results in the employment of the corpus stylistic approach, as a merging of these two disciplines, having the potential to advance the studies of stylistics. This explains how Mahlberg

and Smith (2010) view corpus stylistics as ‘an emerging field that aims to combine questions from literary stylistics with approaches from corpus linguistics’ (Mahlberg & Smith 2010: 450). It is valuable to note here that stylistics does not merely represent the analysis of literature. Moreover, those approaches employed in corpus stylistics can hail from stylistic rather than linguistic domains. This can be exemplified in the analysis of the presentation of speech and thought through corpus, which employs analytical approaches grounded in stylistics as opposed to corpus linguistics.

O’Halloran (2007) considers corpus stylistics to be ‘an infant discipline as evidenced by the relatively small but nevertheless significant work that has taken place’ (O’Halloran 2007: 228), which was reviewed in Wynne (2005), for instance. Although Wynne (2005) states that ‘[i]t is perhaps surprising, then, that ... there is little use of language corpora, or the techniques of corpus linguistics, in the study of literary style’ (Wynne 2005: 1), Mahlberg (2007a: 2) asserts that simultaneously the approaches of corpus linguistics are gaining increasing popularity, **offering the examples of** (Lawson 2000; Adolphs & Carter 2002; Culpeper 2002; Hori 2004; Semino & Short 2004; Stubbs 2005; Adolphs 2006; Scott 2006; Starcke 2006; see also the overview in McEnery et al. 2006: 113–6).

O’Halloran (2012) reports that in recent years, corpus analysis has begun to be employed to support **the** interpretation (and evaluation) of poetry and other literary genres, **giving the** examples of (Adolphs & Carter 2002; Biber 2011; Culpeper 2009; Fischer-Starcke 2010; Hoover 2002; Louw 1993; Mahlberg 2007a; McIntyre & Walker 2010; O’Halloran 2007a; 2007b; Romaine 2010; Stubbs 2005; Toolan 2006). McIntyre (2008) similarly asserts that technological progression in computing and software has enabled the analysis of large volumes of text through corpus stylistics in a manner that was not feasible in the pre-corpus software era (e.g. see Hardy 2004; Semino & Short 2004; Stubbs 2005).

3.4.1 The theoretical concepts and principles of corpus stylistics

The most pertinent principle in respect to corpus stylistics is the assumption of an equivalence between frequency and significance in language data, namely that corpus linguistics assumes particularly frequent features are significant for the discourse structure and meaning of the data. A high frequency of features may function as an opportunity to highlight them in the text, since, because of their frequencies, they assume significance either for the text's content or its structure. Consequently, the analyses in this **thesis** are grounded **in** the most frequent or dominant realisations of the features discussed, that is, **in** the most dominant semantic fields identified **in** a list of keywords of a text and **in** the text's most frequent phrases. Textual meaning, i.e. a text's literary interpretation, is induced from these features, and their collocations and colligations. The fact that the patterns are objective features of the text provides a firm basis to their subjective interpretation, while evidence for the interpretations is gathered by concordance lines for every word or phrase discussed. A further advantage of selecting solely the most frequent items for analysis is that this procedure responds to Fish's (1973) criticism of stylistics, namely that it frequently selects features for analysis which the analyst is convinced are important for the text. The frequency-based approach thus overcomes this subjective choice insofar as possible, and introduces an objective criterion into the analysis.

Mahlberg (2007a) emphasises what she considers to be at the core of the theoretical framework of a corpus, and that which can advance literary stylistics while contributing to its further development:

- 1) language is a social phenomenon;
- 2) meaning and form are associated; and,
- 3) a corpus linguistic description of language prioritises lexis.

(Mahlberg 2007a: 3)

Mahlberg (2007a) continues by defining these three core components through consideration of the nature of language as a social exchange, thus enabling the potential to see meaning as the first purpose of language's use. She describes the second core component as signposting so that through observing repetitive patterns of language use via the corpora we can establish meaning, while those patterns that co-occur with frequency imply the relationship of form and meaning. Mahlberg (2007) **proposes** her third core component as being associated with the linguistic description of meaning resulting from the corpus through establishing the connection between form and meaning, **indicating towards the significance of the lexical component.**

Through the use of corpus stylistic methods, information regarding *wordlists*, *keywords*, *collocations*, *clusters* and *semantic prosody* can be acquired and employed to raise our level of understanding of texts (see Chapter 5 for the discussion of these terms, and then Chapter 6 for their application). Corpus stylistics also offers the opportunity to determine the foregrounding forms that deviate from the linguistic norm, inclusive of the statistical deviation (McIntyre 2010). The use of corpus methodology in investigating Dickens's work is intended to provide a means for objectively assessing Dickens's lexicon. Such analysis can also provide an indication of which aspects of the text are likely to be worthy of further focus (e.g. **see** McIntyre 2010).

Mahlberg and Smith (2010) explain that '[i]t is for the analyst to assist how searches can be narrowed down and complemented with further methodological approaches' (Mahlberg & Smith 2010: 466).

Another clear concept in corpus stylistics is the fact that it begins with the linguistic features of a text in order to explore the impact of those linguistic features on the readers. Widdowson (2008) observes:

In the pre-corpus period ... generally speaking, what directed the selection of linguistic features was some impressionistic sense of literary significance. It worked from the literature to the language. With corpus analysis, however, we have the possibility of working in the other direction.

(Widdowson 2008: 293)

According to Leech (2007), 'the strength of corpus stylistics ... lies in its ability to show stylistic phenomena recurring or developing over whole texts or text collections' (Leech 2007: §2.5). Mahlberg (2010) considers corpus stylistics as being 'the interface of corpus linguistics and literary stylistics', which 'employs methods and approaches of corpus linguistics and links them with concerns in literary stylistics and literary criticism' (Mahlberg 2010: 295). This approach facilitates the arrival at the 'quantification of linguistic phenomena' and the potential 'to compare individual texts against general reference corpora' (Mahlberg 2010: 295). Mahlberg (2010) states firmly that '[s]till, the value of applying corpus methods is defined through the links that can be made between quantitative findings and qualitative analysis' (Mahlberg 2010: 295).

The significant differences between corpus linguistics and corpus stylistics do not lie within the procedural domains of the two subfields, as they both employ similar 'methods and approaches of corpus linguistics' (Mahlberg 2010: 295), but rather they differ in respect to the objectives that the two subfields attempt to realise. There are three main differences that can be identified. Firstly, while corpus linguistics aims to identify the patterns and the norms of the language as it occurs in authentic and natural language use, it is the interest of corpus stylistics, and indeed of stylistics itself, to identify the deviation from such established norms of the language, and then to proceed to explain *the effect of that deviation* in constructing the given meanings of the text. Secondly, corpus stylistics is as keen as stylistics itself to establish a relation between the forms as linguistic choices and meanings that are constructed as a result of those chosen forms of the language. It is not within the objective of corpus linguistics to merely portray a literary work by virtue of a

simplistic quantitative report into contours of texts, but rather the lens of the various corpus tools is employed to realise insights that can be framed ‘against an awareness of the text as a complex, historically realised product’ (Hunt & Carter 2012: 30). Practically speaking, this has resulted in corpus stylistic analyses generally investigating and considering those literary texts highlighted through formal readings and being characterised by the challenges in substantiating them using established means, as opposed to any attempt at producing some kind of comprehensive explanation of a complete literary work (Hunt & Carter 2012). Moreover, those in-depth understandings that are acquired from engaging with corpus stylistics can be employed to enhance and advance current understanding of a narrative in respect to its textual characteristics, while gently drawing both the reader and critic’s attention to the value of the linguistic form in expressing meaning, as opposed to any assertions that the meanings of literary works can be defined through merely ‘decontextualised statistics’ (Hunt & Carter 2012: 30). Milojkovic (2013) suggests that ‘corpus stylistics and corpus stylistics pedagogy has to do with the amount of percentage of occurrences of a specific tendency in meaning’ (Milojkovic 2013b). Thirdly, corpus stylistics is also concerned with those questions raised by literary critics and how they intuitively **respond** to the literary texts, and to subject such interpretations and responses to empirical tests by employing evidence acquired through corpus linguistic methods to establish their validity.

3.4.2 The values and goals of the corpus stylistic approach

Fischer-Starcke (2009) considers that

[t]he benefits of using corpus linguistic analytic techniques in the analysis of literature as opposed to traditional, non-computational approaches are:

- a. the possibility to detect so far unknown meanings of the data through the analysis of so far unrecognized linguistic patterns and
- b. the possibility to base a, by definition subjective, interpretation of a text on objectively existing linguistic patterns.

(Fischer-Starcke 2009: 494)

Semino (2011) clearly underscores that ‘the application of the methods of corpus linguistics to text analysis has made it possible to place the selection and description of stylistically significant features on a firmer empirical footing’ (Semino 2011: 543). From the perspective of Keshabyan-Ivanova (2014: 59), it has been possible through analytical techniques based on **corpora** to (i) empirically measure assertions regarding language use in literature; (ii) source text-based evidence; (iii) draw conclusions regarding the nature of literary and non-literary styles; and (iv) establish significant insight into the form and content of texts.

Carter (2012) considers that ‘[c]orpus stylistics extends practical stylistics and is growing as a methodology within the world of stylistics, linguistics and poetics, enabling more developed and detailed quantitative studies of literary linguistic patterns of meaning formation’ (Carter 2012: 107). Included in the available approaches for achieving this is the ‘use of computer-informed searches of the language of large multi-million word databases, considerably advancing reliability in the identification of the traits of individual authors or groups of authors’ (Carter 2012: 107). Stockwell (2006) states:

The continuities between literary creativity and the creativity apparent in everyday discourse have been revealed in all their complexity largely out of the fruitful interaction of stylistics and corpus linguistics. New methods such as these can be used to explore levels of language from lexical collocations right up to narrative organization.

(Stockwell 2006: 756)

Mahlberg (2007b) indicates towards several practices that corpus stylistics can offer in the analysis of texts, amongst which is ‘the obvious provision of quantitative data’, besides ‘providing various options for the comparison of one text with groups of other texts to identify tendencies, intertextual relationships, or reflections of social and cultural contexts’.

The use of corpus linguistic techniques and strategies', as Carter (2012) suggests, 'allows significant linguistic patterns to be identified that would not normally be discernible by human intuition, at least not over the extent of a whole novel or long narrative poems and dramas' (Carter 2012: 107–8).

In his introduction to Mahlberg (2007c: Chap. 2), McCarthy (2007: 19) states:

With texts stored as a corpus, scholars can access frequency lists and concordances from the works of an author or groups of authors. These in turn can be compared statistically with texts by other authors, or with larger benchmark corpora, in ways that offer the promise of isolating authorial 'fingerprints'.

(Mahlberg 2007c: 19)

An example of the systematicity that corpus stylistics provides would be that which Mahlberg (2009) argues for in terms of how 'Dickens often references the ways that characters use household objects such as a watering-pot or a knife and fork as a way of drawing attention to their emotional states' (Litosseliti 2010: 101). Moreover, Mahlberg 2014 (380) suggests that '[c]ounting and comparing is also important to corpus stylistic work' by placing 'greater emphasis on a qualitative dimension than computational stylistics does. It is exactly the counting and comparing that contributes the additional systematicity to literary stylistics which seems to make corpus stylistics such an attractive undertaking'.

In addition to the systematicity, which refers to the account of a text systematically, in which several linguistic features are associated with their potential effects, corpus stylistics can further provide considerable evidence in terms of quantity, besides the 'systematic evidence' for unavoidable subjective interpretation (Mahlberg 2009: 48). Corpus stylistics, as Mahlberg (2009) suggests, can demonstrate the connection between 'literary stylistics and corpus linguistics' (Mahlberg 2009: 48). Corpus linguistics is concerned with 'the description of linguistic norms and deviations from these norms', whereas literary stylistics intends to identify any

‘specific linguistic phenomena across different sets of texts’, while also ‘[raising] our awareness about the need for interpretation in corpus work’ (Mahlberg 2009: 48). Corpus linguistics can approach and analyse literary texts from different perspectives depending on the targeted linguistic feature and the adopted methods, with case studies having focused on collocations, clusters, keywords and key semantic domains (Mahlberg 2009). Fischer-Starcke (2010) states that the ‘potential and goals of corpus stylistic analyses [...] allow us to develop analytic techniques for investigating various research questions, to evaluate the success of different research techniques for different sets of data, and to gain new literary and structural insights into the data’ (Fischer-Starcke 2010: 10).

Carter (2010) meanwhile considers the embracing of corpus stylistics in textual interpretation and understanding as ‘empowering the reader with an equipment that ... allows them to take it further for themselves’ (Carter 2010: 65–7). Moreover, he considers employing such methodologies ‘to advance the study of style’ and ‘to better account for the processes of meaning construction which are the basis for our understanding and interpretation of text’ (Carter 2010: 65–67). This is due to stylistics’ concerns with meaning construction, which is considered to be the manner in which readers understand and interpret a text. Yet, the selection and use of corpus stylistics should empirically aim to realise the process of constructing the meanings of a text, although this by no means implies that there would not be different readings for the same texts due to the differing inherent factors which impact on the reading process (e.g. background, culture and language proficiency).

In their study of William Blake’s poem collection *Songs of Innocence and Experience*, McIntyre and Walker (2010) state that ‘corpus analysis validates some of the subjective critical responses’ to texts while explaining how ‘[q]uantitative analysis guides qualitative analysis’ (McIntyre & Walker 2010: 522). The employment of corpus stylistics facilitates greater understanding regarding word use and how it functions in context. This further supports conducting an effective

qualitative analysis, since the focus is based on context and use. For example, for a specific word it would be feasible to discover how it was used in a given text or throughout an entire work (e.g. Dickens's fiction and non-fiction work). What becomes pertinent in respect to the use of corpus stylistics is that the systematicity in this study, as well as the comprehensiveness of Dickens's works, facilitates the treatment and dealing with several features simultaneously that contribute directly to creating the meanings of the text. The analysis of Dickens's corpus can thus result in the achievement of one of corpus linguistics' significant goals in general, which is to form a number of generalisations on the linguistic patterns of Dickens's texts as data, or Dickens's language as an author. This should answer the question of whether Dickens has systematicity in his semantic level, word choice, phraseology, collocations and thoughts.

A consideration of the range of studies that harness corpus stylistic approaches include Louw's (1997) investigation of 'semantic prosody'; Hoover (1999) analysing quantitatively the language of Golding's *The Inheritors*; Culpeper (2002) examining the 'key words' of the main characters in *Romeo and Juliet*; Mahlberg (2007) exploring 'clusters' in Dickens's style; and Hoover (2002) relating the most frequent words and clusters to their authors. McIntyre and Walker (2010) assert that 'a corpus stylistic approach can be used to validate some of the more subjective comments of film critics, and to provide insights into the linguistic construction of particular genres' (McIntyre & Walker 2010: 522). Needless to say, stylisticians are 'not advocating a "method" for mechanically generating an interpretation of' texts, as O'Halloran (2012: 173) notes (see McIntyre (2014: 392-3) for an up-to-date review of work entailing the use of corpora in stylistic studies).

One of the values of employing corpus stylistic methodology lies in its ability to facilitate the stylistic analysis of long literary texts, since it is virtually impossible to analyse 'a long novel in close stylistic detail [which] could take a lifetime' (Short 1996: 255). The methods utilised in corpus stylistics are adopted to promote a

'complementing interpretation' to the intuition and subjective interpretation, and detailed tools for the stylistic analysis (Mahlberg & Smith 2010: 450). O'Halloran (2007) asserts that '[t]he value of the corpus stylistic approach ... is that it can help reveal such subtleties that even some close readings may not detect' (O'Halloran 2007: 241). There are numerous corpus linguistic methods that can be implemented to satisfy the goals of stylistics. Thus, the term 'corpus stylistics' implies that corpus linguistic methods are of benefit in facilitating the linguistic interpretation of texts. Despite the expediency of corpus linguistic methods, the fact that literary analysis cannot be computationally programmed rather emphasises the fact that the human interpreter is fundamental to the process of interpretation, since those tools are to be exploited to assist the interpreter in achieving his/her goals. Therefore, the practicality lies in 'corpus methodologies that can support the stylistic analysis' (Mahlberg & Smith 2010: 450). Corpus methodologies in this regard include, among others, wordlists, keyword lists, concordances, intratextual comparisons, cluster lists and distinctive patterns of language preferences (Mahlberg & Smith 2010; Römer & Wulff 2010), as will be discussed in Chapter 5. The use of these methodologies is justified in that 'the value of applying such corpus methods depends on the links that can be made between quantitative data and thematic arguments and interpretations' (Mahlberg & Smith 2010: 49–50), the difference between corpus methodology and corpus methods being that methodology can be taken in this context as a group of methods that are governed by theoretical concepts (methodology) in order to respond to a set of stylistic questions.

While employing corpus **methods** in stylistic analysis, the '[c]reative features of a literary text that could be of great interest to a stylistician' (Mahlberg 2009: 48) should be taken into consideration, lest they 'disappear below the waves' of general corpus (Sinclair 2007: 3). For such 'creative features', the rational process and intuition are demanded, which simultaneously reveals that utilising corpus methods does not transform stylistic analysis into a mechanical process as the data and

findings cannot speak for themselves, rather they need to be interpreted and made meaningful. Corpus stylistics can be viewed as a new discipline, where literary texts receive greater attention as data is compared with other genres which implement corpus linguistic tools to stylistically analyse a text.

Corpus stylistics thus tends to avoid generalisation about language, but rather determines a phenomenon within the literary texts. Consequently, if a generalisation is to be made or claimed, it has to be assumed only within that given literary text; and the more specific the corpora (e.g. Dickens's work), the greater the generalisation is limited to such texts, even if it can extend beyond these confines on certain occasions due to the similarities in genre that occasionally occur.

Corpus stylistic tools are intended to identify the distinctive linguistic patterns of literary texts. Determining and describing the linguistic patterns of Dickens's work through corpus can be reliable in terms of textual analysis. The size of texts is still limited when compared with general linguistic data, where corpora are in the millions of words, to deduce generalisations associated to literary texts. This is justified due to the fact that this study is solely concerned with Dickens's work, and any such linguistic patterns can be basically ascribed to this described text. Hence, when a claim is made that some features are common, based on the investigated work, this would make sense. Thus, the use of corpus tools in stylistic analysis differs in terms of aim and application from the use of the same techniques employed for general large corpora (Mahlberg & Smith 2010: 49–50). The justification for investigating a small corpus (i.e. Dickens's) is to identify the uniqueness of such data in terms of what is uncommon about the language that has a consequential effect on the language style.

The application of these different approaches employed through corpus linguistics to study and analyse literary texts is intended to offer tools to interpret

the text and validate, or in some cases invalidate, the intuitively subjective accounts of literary texts. The usefulness of applying corpus stylistic methods includes

the application of quantitative methods to literary texts ... the reflection on the types of questions we can ask and attempt to answer ... [c]orpus stylistics can add to the analysis of a short text extract ... [looking at] a wider textual perspective that draws on clusters as pointers to local textual functions and building blocks of textual worlds.

(Mahlberg 2009: 48)

Fischer-Starcke (2010) considers corpus stylistics as being ‘the linguistic analysis of electronically stored literary texts’, as a combination of stylistics – ‘the linguistic analysis of literary texts’ – and corpus linguistics – ‘the electronic analysis of language data’ (Fischer-Starcke 2010: 1–2). She states that corpus stylistics makes it possible:

1. to study how meaning is encoded in language and to develop appropriate working techniques to decode those meanings, and
2. to study the literary meanings of texts.

(Fischer-Starcke 2010: 1–2)

This is not to suggest that it is unfeasible to interpret a literary text without employing corpus stylistic methods, but rather it results in acquiring greater insights of large corpus such as in the case of the DCC, which will be demonstrated in Chapter 6.

The rationale that leads **me** to opt for the use of corpus stylistics is that it combines both ‘the use of corpus linguistic techniques and the goals of stylistics [that] complement each other as both disciplines decode linguistic patterns and their meanings in texts’ (Fischer-Starcke 2010: 1). By adopting this approach, this study intends to define one of the approaches to the analysis of the meanings that are encoded in the literary texts.

This combination of the two disciplines also generates an analytic potential of corpus stylistics, allowing for the decoding of the meanings of literary texts that cannot be detected either by intuitive techniques (as in literary studies) or with the necessary restriction to short texts or text extracts (as in traditional stylistics). 'Corpus linguistic techniques allow (1) a systematic and detailed analysis of large quantities of language data for lexical and/or grammatical patterns and (2) to subsequently decode the meanings of these patterns' (Fischer-Starcke 2010: 1).

One of the **strengths** of corpus stylistics can be understood in terms of how flexibly it combines different **methods** in order to address the research questions, especially those that are stylistically related to literary analysis. A further related point is that the research questions raised in this study dictate how the tools should reveal the information so that they can **comprehensively** respond to the inquiry's standing questions. Otherwise, such questions will be restricted to those which the corpus tools can offer, exclusive of belief in the potential to control the tools to meet or respond to the required answer. In essence, we are to control the corpus tools so that they answer our questions, rather than such tools assuming control to thus respond to those questions that can only be answered mechanically.

Murphy (2007), **quoting Biber et al. (1998)**, highlights **that**:

[T]he underlying characteristics of a corpus-based approach to language study are that it is empirical, it uses a large corpus of natural texts collected on a principled basis, computer and manual analyses of the corpus are carried out, and both quantitative and qualitative techniques of analysis are employed. The consequent strengths of such an analysis lie both in its scope and reliability.

(Murphy 2007: 67)

Meanwhile, Stubbs (2005, cited in Murphy 2007: 67) **underlines** a further advantage of utilising computer software as being its ability to highlight textual features of potential significance that may have been overlooked by literary critics.

Corpus stylistics also reveal 'how empirical corpus evidence can usefully provide substantiation of such initial evaluations of literary works' (O'Halloran 2007: 33).

O'Halloran (2007) claims:

Work in corpus stylistics has on the whole focussed on showing the power of large corpora in providing a systematic description of a literary work's salient features ... or assisting in the interpretation of a literary work ... that is, what the literary text means to an individual stylistician after some reflection. There has been some focus, too, on how corpora might provide evidence of underlying cognitive processes.

(O'Halloran 2007: 35)

O'Halloran (2007) continues by mentioning that 'formulaic sequences can include both collocation and phraseology' (O'Halloran 2007: 36) **within its broad definition.**

According to Keshabyan-Ivanova (2014), '[c]orpus-based analytical techniques' have made it 'possible to test empirically claims about the language of literature, to search for and provide evidence from texts, to establish the norms of literary and non-literary style, and to have in-depth insights into the texts' structures and meanings' (Keshabyan-Ivanova 2014: 59). To summarise, corpus stylistics values are both objective and reliable.

3.4.3 The difference between corpus linguistics and corpus stylistics

Despite the tendency of the 'focus of interest' of both literary stylistics and corpus linguistics to be different, as Mahlberg (2007b) observes, stylistics focuses on the 'distinctive' features of the text 'and it investigates deviations from linguistic norms that trigger artistic effects and reflect creative ways of using language' (Mahlberg 2007b: 221). On the other hand, '[c]orpus linguistic ... mainly focuses on repeated and typical uses that do not only hold in one text, but are found across a number of texts in a corpus' (Mahlberg 2007b: 221), which reflects a typical feature of corpus linguistic studies that aspires towards generalisations regarding the use of a

language. In explaining the 'link between corpus linguistics and literary stylistics', Mahlberg (2007b: 221) highlights the following two points. The first is that '[b]oth are interested in the relationship between meaning and form', and secondly that '[s]tylistics puts an emphasis on how we say what we say and corpus linguistics also claims that what we say depends on form, i.e. the patterns which are attested in corpora' (Mahlberg 2007b: 221). Hunt and Carter (2011) argue that corpus linguistics as a methodology used in the analysis of texts still

has two defining characteristics: firstly, analysis is conducted on large collections of digitally stored, naturally occurring language data (corpora), which range from several thousand to over a billion words. Secondly, in order to systematically investigate such a vast volume of language, corpus-based research employs computational software to capture textual patterns that escape attention during normal reading and to organise the data for manual reading in previously impossible formats.

(Hunt & Carter 2011: 29)

Wynne (2005) underscores that both stylistics and corpus linguistics are *empirical approaches* concerned with a linguistic description grounded in the evidence of how language is employed through collection and analysis in corpora. Mahlberg and McIntyre (2011) consider that 'Louw's [1993] research provides convincing evidence of the potential for corpus linguistic methodologies to benefit stylistic analysis ...' (Mahlberg & McIntyre 2011: 205).

The contrast between corpus linguistics and corpus stylistics is presented by Carter (2010) in respect to the application, focus and priority. Corpus linguistics on the one hand places priority on lexis, considers what lexical patterns can be identified with the aid of computer software, and locates words that co-occur with frequency, while placing particular emphasis on collocation and dismissing the 'idiosyncrasies of individual texts'; while on the other hand, the lens of corpus stylistics focuses on the specific qualities that an individual text may have, due to stylistics paying closer attention to the 'deviations from linguistic norms' that lead to

the manifestation of 'artistic effects' (Carter 2010: 65–7). Methodologically, corpus stylistics will search the 'lexical population' for that which is considered significant; while analytically, the procedure quantitatively assesses the statistical relevance that is acquired 'from a corpus-informed count' (Carter 2010: 65–7). Nevertheless, there is also the involvement of the analyst in making 'qualitative decisions and interpretive acts' in anticipation of and response to the emerging results, with the phase of interpretation being relatively subjective in nature (Carter 2010: 65–7).

Finally, Wales (2006) confirms 'the fact that primacy [in stylistic analysis] is assigned to language', which thus invalidates the claim that stylistic analysis is 'simply a mechanistic description of the formal features of texts for their own sake' by arguing that such analysis aims 'to find linguistic evidence for a critical judgment; to ground intuitions or hypotheses in a *rigorous, methodical, and explicit textual basis*; to produce an analysis that is verifiable' (Wales 2006: 213; emphasis added). Thus, when studying Dickens's lexicon to facilitate the understanding of non-native readers, such study should be rigorous, with a methodical process that is explicit to the reader. A further step is to describe the relationship between this lexical phenomenon and comprehension, and then draw a virtual image of the real world based on the text. Thus, the use of corpus stylistics is intended to satisfy this study and its claims for empirical treatment and verification.

3.4.4 Using the corpus stylistic approach in teaching literature

This concept leads on to the idea of advancing the teaching of literature from a subjectivity-based approach to one based on objectivity. In this regard, as the interpretations made by critics are grounded in intuitive procedures and impressionistic feelings, there is a necessity for greater objective interpretation through the support of corpus linguistics techniques. By the same analogy, teaching literature to non-native readers should be based on linguistic concepts by means of corpus linguistic analyses, rather than merely on pedagogical or educational

intuition where the suggested method is to 'talk about' and 'repeat', and then a certificate is granted for 'understanding the text'. Introducing linguistic insights into the teaching would thus be of use if literature assists in making such a process relatively systematic, measurable and evidence-based. More interesting is that the linguistic insights are extracted from the texts themselves, which are being taught to ensure that the target text has been approached in a linguistic, systemic and relatively objective manner.

The **justification** for utilising corpus stylistic techniques in this thesis lies in the fact that they will facilitate the systematic study of the linguistic features related to the construction of meanings in Dickens's works (e.g. lexis), and how these can be introduced to non-native speakers to promote acquisition of the basic knowledge of meanings. This approach will be established based on solid, non-subjective evidence elicited by means of corpus **methods** that result in at least the minimum level of understanding necessary before progressing further into deeper reading and understanding of Dickens's literary texts.

The use of corpus stylistics to facilitate the teaching **of** literature, such as Dickens's work for non-native readers, offers linguistic/stylistic insight, as opposed to that which is **traditional and subjective**. In order to suggest an appropriate analogy, since stylistic studies inoculate literary criticism with objectivity, the goal is to enhance objectivity in teaching literature by means of corpus stylistics, and to reduce the subjectivity that might be found in pedagogical studies in relation to the reading or teaching of literary texts. With regards to what corpus stylistics can add to the analysis of texts, the primary focus in stylistic analysis is on the linguistic phenomena within which meanings are encoded. With the benefit of technology, linguistic features can be decoded through corpus methods, which will thus assist in determining the linguistic features that are sought to address some of the questions of stylistics (Simpson 2004: 2; Trask 1999, 2007: 280; Fischer-Starcke 2010: 6). There is **a** variety of software (e.g. see Short 1996; McIntyre 2010: 180–2) that has been

developed to enable the identification of linguistic phenomena and features, and results in the analysis of large collections of texts being possible, feasible and achievable.

It should be noted that the claim that meanings in texts can be triggered by specific linguistic features that can be determined through corpus linguistics does not restrict that trigger to only those specific features. Rather, this is one of the features that corpus linguistics can offer to assist in directing the focus towards the salient features that it might benefit learners or readers to focus on. As Widdowson (2008) states, 'one can acknowledge that the value of corpus analysis is that it can provide textual substantiation to impressionistic interpretation' (Widdowson 2008: 294).

Reppen (2009) provides a number of examples to demonstrate how corpora can inform the teaching of the English language. She discusses **the prominent linguistic characteristics as well as the manner in which huge volumes of realistic data can be facilitated for teachers to engage with when developing classroom exercises**. Examples of such salient linguistic features include the 'features of spoken language students will frequently hear outside the classroom, or what grammatical features students will encounter in the different types of texts that they will be reading or writing' (Reppen 2009: 207–8). The focus in this study will be placed on the linguistic patterns of Dickens's work that trigger the comprehension of the text by non-native readers, as corpus stylistics can assist English language teachers and learners in focusing on Dickens's linguistic patterns in order to raise awareness about Dickens's style. The significance of identifying the linguistic features in Dickens's work lies in two main areas. The first relates to the decision of what and how to teach, as suggested by Reppen (2009). Such choices are established in relation to the 'systematicity' (Mahlberg 2009: 48) that corpus stylistics provides.

It has been argued by Kettemann and Marko (2004) that engaging with corpus analysis to examine literary texts offers considerable potential support to language

students. The authors postulate that the application of concordance can benefit such students in their text-based investigations, either before or following their initial reading of the text. It is further argued that language awareness will be elevated, and moreover the insights of how certain linguistic structures can facilitate understanding of literary texts, while additionally consolidating discourse awareness by affording greater discrimination between texts of a literary and non-literary nature. Kettemann and Marko (2004) further assert that engaging with literary texts through corpus analysis could prove to be of value for all aspects of teaching.

With respect to the application of corpus stylistics in teaching literature, Hardy (2007) finds:

Concordancing programs allow instructors to demonstrate not only the discovery of the broad statistically dominant patterns that create intuitive generalizations about texts but also the discovery of the almost endless small-scale variations in those dominant patterns, variations that illustrate the true complexity of the text and of stylistics.

(Hardy 2007: 88)

Louw (2007) also **believes** that ‘both teacher and learner become united in a single endeavour: the scientifically respectable pursuit of meaning by and through the nascent discipline of collocation as instrumentation for language’ (Louw 2007: 104). This is associated with an approach that depends considerably on the role of collocations in order to define how the meanings of a text can be constructed. Additional available approaches include the study of ‘a word, lemma, multiword expression or a grammatical construction [syntax]’ (Paquot 2010: 29), as well as text-grammar, phonology, semantics and pragmatics (Fischer-Starcke 2010: 41).

3.4.5 Limitations of the corpus stylistic approach

Semino (2011) indicates the limitations of applying the corpus stylistic approach in stylistic analysis by stating:

It needs to be acknowledged that corpus-based and computer-aided methods have inevitable limitations, both in terms of what can be investigated and how. Hence, these methods cannot replace the more traditional, intensive, approach to stylistic analysis, nor do away with the role of the analyst's intuitions (which are, for example, involved in the interpretation of the output of software tools). Nonetheless, these methods are an invaluable addition to the methodologies available to stylisticians, as they enable scholars to test out empirically their intuitions about texts, and to study patterns that could not be realistically investigated manually.

(Semino 2011: 547–8)

Kettemann and Marko (2004) nevertheless assert caution regarding the limitations of engaging with stylistics through the corpus-based approach, namely (i) the inherent need to disassemble texts into resulting fragments that by definition dispose of the completeness and integrity of the original text, and therefore remove the potential for interpretation through linear reading; (ii) the emphasis on the superficial surface of the texts at the expense of critical and deep reading, which erodes the ability to attain in-depth meaning and critical engagement with the texts; and (iii) the obfuscation of literary concerns due to the ability of corpus analysis to engage linguistically, at the expense of being able to respond to more profound literary enquiry. However, the flaw in these concerns is the assumption of corpus analysis being viewed as the sole instrument of analysis for considering a literary text, or that it can represent a direct replacement for conventional deep and linear reading (Kettemann & Marko 2004; see also Mahlberg 2007b for the importance of applying intuition to the analysis besides interpreting the quantitative data gained by the corpus stylistic approach).

Hamilton (2008) finds that '[e]ven defenders of corpus stylistics openly admit that "the heart of stylistic analysis" is qualitative rather than quantitative (Semino & Short 2004: 7), which means statistical data are perhaps most useful when they augment rather than replace qualitative analyses' (Hamilton 2008: 560).

3.5 Conclusion

There are a number of possible limitations to the use of corpus stylistics 'in the case of researchers who may automatically apply quantitative searches for distinctive linguistic features without reading literary works closely and sensitively' (Hori 2004: 207), producing surface interpretation without grasping the meanings in a substantial manner. A further concern is related to generating a keyword list of Dickens's work. While such a list can only be accurately representative of Dickens's work, it can also be exploited as an authentic example of the common lexicons of a given period of time, e.g. the Victorian era.

Among the challenges that this methodology will attempt to address is the issue of 'lemmas' and 'related forms' (Scott 2010: 51–2). With respect to lemmas, corpus software can lemmatise lexemes by '[grouping] together varying words or forms of words' (Matthews 2007: 220). Thus, the two terms 'headword' or 'lexeme' can be considered interchangeable when applied to vocabulary or keyword analysis: they are both in their basic forms that can extend to other possible forms. Further potential terms that can represent the same phenomena include the *canonical form* and *dictionary form*.

It should be clarified that the purpose of utilising corpus stylistics in facilitating Dickens's work in this study is primarily to foster elementary reading and pave the way for readers to engage more fully with the text, with the next stage being the intention to explore more deeply the readings of the same texts. This is because, beginning with the very initial step in reading, this study will be concerned with the elementary or primary readings of Dickens's texts, i.e. comprehending the basic meaning of the text, after which learners can be expected to be equipped with the required lexical knowledge to enable attempts at exploring the text in different manners and readings, as Mahlberg (2009) explores in *The Pickwick Papers*.

One of the limitations of the corpus stylistic approach is nevertheless that it provides vast quantities of information that are **challenging** to handle for stylistic analysis, with Widdowson (2008) cautioning that '[t]he problem of relevance, however, remains, and indeed becomes more difficult precisely because we have so much linguistic information to deal with. How do we decide what to select as significant?' (Widdowson 2008: 294). The question of how corpus linguistic techniques can improve on manual analytical methods will be explained in detail in the Chapter 6: **the** DCC analysis.

Chapter 4 The Charles Dickens Complete Corpus

4.1 Introduction

This chapter aims firstly to present the process of constructing the Charles Dickens Complete Corpus (hereafter referred to as the DCC) through observing the standards of corpus linguistics. Secondly, it will offer a detailed description of the DCC's contents.

In order to address the research questions on facilitating engagement with the works of Dickens for non-native readers, constructing a corpus encompassing the full texts of the complete works of Dickens became a prerequisite and an essential platform prior to investigating the texts. The DCC was compiled in accordance with the theoretical guidelines of corpus linguistics standards. In this chapter, those decisions taken in constructing the DCC will be documented so as to precisely describe the procedures through which the compilation of this corpus has progressed.

The motive for creating this DCC is the lack of a pre-existing complete corpus for Dickens's works, as no complete corpus has been located that **includes**, literally, the 'complete' works of Dickens. Furthermore, despite the existence of two corpora (entitled **the** 'Charles Dickens Corpus' and 'Dickens Corpus', respectively), neither contain the author's complete works. The first is available from the English Department in Giessen University, Germany. By appraising the contents of this corpus, it was found to contain a less than all-inclusive range of Dickens's works, as they will be explained later in this chapter. Nevertheless, the contents of the Giessen University's Dickens **Corpus** include forty-three documents, while the total number of word tokens

is 4,621,685. The contents are presented, as listed in the corpus, in Table 4.1 below:

Table 4.1 The Dickens Corpus of the University of Giessen: 43 items

Holiday Romance
Hunted Down
To Be Read at Dusk
The Haunted Man and the Ghost's Bargain
The Battle of Life
The Chimes
A Christmas Carol
The Cricket on the Hearth
Three Ghost Stories
A Message from the Sea
Doctor Marigold
Going into Society
Mrs. Lirriper's Legacy
Mugby Junction
No Thoroughfare
Somebody's Luggage
The Holly Tree
The Lazy Tour of Idle Apprentices
The Perils of Certain English Prisoners
The Seven Poor Travellers
The Wreck of the Golden Mary
Tom Tiddler's Ground
Mudfog and Other Sketches
The Loving Ballad of Lord Bateman
Master Humphrey's Clock
The Pickwick Papers
Oliver Twist
Nicholas Nickleby
The Old Curiosity Shop
Barnaby Rudge
Martin Chuzzlewit
Dombey and Son
David Copperfield
Bleak House
Hard Times
Little Dorrit
A Tale of Two Cities

Great Expectations
Our Mutual Friend
The Mystery of Edwin Drood
The Uncommercial Traveller
The Lamplighter and Some Christmas Stories

The other corpus for Dickens’s works is the ‘DCorp’ by Mahlberg (2013), consisting of twenty-three texts, which includes all fifteen of his novels (presented in italics in Table 4.2), seven stories and sketches and ‘one text that can be classified as non-fiction (“American Notes for General Circulations”)’ (Mahlberg 2013: 42–3). The DCorp comprises approximately 4.5 million words, while the list of contents can be found in Table 4.2 below.

Table 4.2 The DCorp’s contents: 23 items

American Notes for General Circulation
Bleak House
The Battle of Life
Barnaby Rudge
A Christmas Carol
The Cricket on the Hearth
The Chime
David Copperfield
Dombey and Son
The Mystery of Edwin Drood
Great Expectations
Hard Times
The Haunted Man and the Ghost’s Bargain
Little Dorrit
Martin Chuzzlewit
Nicholas Nickleby
The Old Curiosity Shop
Oliver Twist
Our Mutual Friend
The Pickwick Papers
Sketches by Boz
A Tale of Two Cities
The Uncommercial Traveller

Since this study aims to investigate the complete works of Dickens, these two corpora do not suit the needs of this study in several respects. The first reason is that they do not collect the complete works of Dickens, though none of them claimed so. Secondly, apart from excluding some entire work or collections of Dickens such as *A Child History of England*, *Reprinted Pieces*, *Sunday under Three Heads*, *Miscellaneous Papers*, *Plays and Poems* and *Letters and Speeches*, some of the documents included are partially extracted from other works and being merely essays, short stories, or books/chapters in larger works. *Hunted Down* and *To Be Read at Dusk*, for instance, in Giessen University's Dickens corpus are respectively an essay and short story extracted from *Reprinted Pieces* and consequently are not complete works by themselves. Table 4.3 demonstrates in more detail solely the extracted-from-other-works contents of Giessen University's Dickens corpus. In the case of the DCorp by Mahlberg (2013), the five items of *The Battle of Life*, *A Christmas Carol*, *The Cricket on the Hearth*, *The Chime* and *The Haunted Man and the Ghost's Bargain* refer essentially to one volume entitled by Dickens himself as *Christmas Books* (Davis 1999: 65), albeit being printed individually in occasions. Thirdly, the DCC attempted to maintain the relative balance expected in specialised corpora by including other text types which were not included in the other two Dickens's copra. *A Childs History of England* and *Miscellaneous Papers*, typically classified as nonfiction works, were included in the DCC, besides Dickens's letters and speeches.

Table 4.3 Contents of the University of Giessen’s Dickens Corpus extracted from other works

Work	Description and notes
Hunted Down	A short story in <i>Reprinted Pieces & Sunday under Three Heads, and Other Tales, Sketches, Articles, etc.</i> vol. xxxiv in National Edition
To Be Read At Dusk	An essay from <i>Reprinted Pieces</i> , vol. xxxiv in National Edition
A Christmas Carol, The Chimes, The Cricket on the Hearth, The Battle of Life, The Haunted Man and The Ghost’s Bargain	These are the five books/stories collected as <i>Christmas Books</i> vol. xvi in National Edition
Three Ghost Stories	Three stories; <i>The Signalman, The Haunted House</i> and <i>The Trial For Murder</i> all are extracted from <i>Christmas Stories</i>
A Message from the Sea, Doctor Marigold, Going Into Society, Mrs. Lirriper’s Legacy, Mugby Junction, No Thoroughfare, Somebody’s Luggage, The Holly Tree, The Lazy Tour of Idle Apprentices, The Perils of Certain English Prisoners, The Seven Poor Travellers, The Wreck of the Golden Mary and Tom Tiddler's Ground.	All these documents are extracted from <i>Christmas Stories</i>
Mudfog and Other Sketches	This is a collection of Dickens's early sketches found in <i>Sketches by Boz II</i> – Illustrative of every-day life and every-day people – <i>Sketches of Young Gentlemen</i> – <i>Sketches of Young Couples</i> – <i>The Mudfog Papers</i> in vol. ii National Edition
The Lamplighter	A story in <i>Reprinted Pieces</i> , vol. xxxiv in National Edition

Here are two points to note about two items in the Giessen University's Dickens Corpus. The first one is *Some Christmas Stories*. There is no such work by Dickens entitled 'Some Christmas Stories', but rather it contains six Christmas stories, namely *A Christmas Tree*, *What Christmas is as we Grow Older*, *The Poor Relation's Story*, *The Child's Story*, *The Schoolboy's Story* and *Nobody's Story* that appeared in *Christmas Stories from 'Household Words' and 'All the Year Round' vol. I*, National Edition.

The second document in the Giessen University's Dickens Corpus is *The Loving Ballad of Lord Bateman*. The inclusion of this document in the corpus seems to be based on the assumption that it was composed by Dickens. Tracking this work casts some doubt on its authorship, with Davis (1999) stating that it is a '[c]lockney version of a traditional ballad, illustrated by CRUIKSHANK and published by Charles Tilt in 1849. Dickens wrote a preface and some notes for the volume' (Davis 1999: 220). Shattock (2000) also confirmed that 'Dickens wrote the preface and notes and adapted at least part of the text, based on a traditional ballad. Rptd by Charles Dickens and W. M. Thackeray' 1969, with note by L. C. Staples' (Shattock 2000: 1245). Likewise, Jordan (2001) explained that 'because [Dickens] was contracted to Bentley not to write for any other publisher – he wrote the Preface and facetious notes for George Cruikshank's rendering of a popular folk song, *The Loving Ballad of Lord Bateman* (1839)' (Jordan 2001: 44). Furthermore, I did not find this work included in the National Edition of Dickens's Complete Works, being of the opinion that it was mainly a work composed by William Makepeace Thackeray, not Dickens. Therefore, there would be no sound reason to include it within Dickens's works.

For those reasons described above, the need to construct the DCC arises due to the absence of a comprehensive corpus. In order to avoid

deficiency in determining the complete works of Dickens, the first step was to precisely establish what works should be included in the proposed DCC.

4.2 Collecting Dickens's works

Due to the proliferation of Dickens's works in different formats, it was uncomplicated to collect the DCC texts as they all exist in electronic format. However, these were dispersed over numerous websites and various university library archives, which necessitated on occasion their tracking and detailed revision before their final inclusion in the corpus body of the DCC. It was less onerous to exploit the availability of electronic texts of Dickens's works as the hosting websites; for example, the Project Gutenberg and the University of Adelaide library contain the full-text content of his novels, novellas, short stories and journal articles, letters and speeches. Dickens's works in these websites are typically stored as individual texts while also being searchable; thus enabling the reader to obtain the work of a single author or a specific work.

Through searching for electronic texts of Dickens's works, the most relevant internet archive was found to be Project Gutenberg: a digital library containing over forty-five thousand documents. The collection of these free items, either due to being post-copyright or where permission has been granted by the copyright owner, comprises the full texts of public domain books. The documents are primarily available in TXT (plain text) format, while other formats may be available including HTML (hypertext markup language), EPUB (electronic publication) and PDF (portable document format).

Project Gutenberg, deemed as the most relevant internet archive, was selected as the primary source to download the texts of Dickens's works, while other resources such as the University of Toronto's Robarts Library,

Trove (the National Library of Australia) and the library of Harvard University have **also** been used. As for the other internet archives that have been utilised besides Project Gutenberg, these still offer some texts of Dickens's work in different formats. The texts have been primarily scanned from the printed copies, and then saved directly to PDF files. From the PDF version, they can be automatically converted into either TXT or HTML files by the website itself, without any requirement for recognition software to identify the texts written on the scanned papers. Unfortunately, the output TXT files would not be in the same quality as Project Gutenberg or the University of Adelaide University library in terms of the proofreading. Although the accuracy of the text converted from the scanned pages into electronic format is remarkable, proofreading is still necessary as letters and numbers can be misidentified. Several examples that were noted include: *I'll* instead of *I'll*, *I' ve* instead of *I've* and the use of American spellings on occasion, where the hardcopy employs British English.

Kennedy (1998) 'reports that current scanners have challenges in identifying hyphens, apostrophes and certain letters or groups of letters such as *a* (car is rendered as cor), *cl* (clear becomes dear), *in* (innate becomes mnate) and the number *1* vs. the letter *l*' (Kennedy 1998, cited in Baker 2006: 34; italics added). I found this indeed to be true following my extensive revision of texts obtained from such archives utilising the word processor to identify the misspelled words. These cases of misidentified and misspelled texts required proofreading, which I have conducted for some of the less common works of Dickens. Due to the advantages of the internet, there was no need to manually type full document reproductions of each of Dickens's work included in the DCC; however, the process did require 'copying', 'pasting' and editing to ensure that some of the documents matched their printed versions of Dickens's works, in addition to inserting some letters and speeches manually

as they could not be located in an electronic format after conducting online searches. The texts that **were** added manually are presented in Appendix 4.1.

When saving the TXT files of Dickens's works, it was attempted to preserve the language data as accurately as possible to mirror the actual printed copies. In order to achieve this accuracy, great effort was undertaken to maintain all the data found in the texts, inclusive of speech marks or accented characters, since Dickens used other characters besides English, such as those found in French or Italian names and vocabulary describing his journeys and the locations he visited, especially in his letters. In order to preserve Dickens's language data that occur in Latin alphabets with diacritics and orthographic ligatures, for example Æ and æ, the TXT files were saved through a 'Unicode' encoding system, that is, the '[i]nternational character-encoding system designed to support the electronic interchange, processing, and display of the written texts of the diverse languages of the modern and classical world' (Britannica Concise Encyclopaedia).

Following collection of the electronic data of Dickens's works, these needed to be spell-checked and corrected for errors. The aim was to match as accurately as possible the printed versions in hand, with the process being to verify the first line of each paragraph to ensure that the text matched the printed copy. **On** rare occasions, I would look for an electronic version of the text other than Project Gutenberg to find an identical text to the printed copy. This happened with the novel *Great Expectations*. I also checked the opening and closing parts of each chapter, as well as ensuring compliance with British English spelling conventions via the word processor spell-checker. This helped in identifying American spellings where the archive from which the texts had been sourced was from American universities. While Project Gutenberg affords valuable opportunities for selecting and downloading the texts in various formats, such documents needed to be revised and lightly

edited so as to match the printed volumes of Dickens's works. Those hard copies from which the DCC is based are available for later reference if this corpus is to be developed further in future.

4.3 The retitled works of Dickens

One of the challenges encountered while collecting Dickens's work is that the number of works ascribed to him are considerable; nonetheless, they are not all titled by the author as a number of them have merely been extracted from other works, as shown above. Therefore, on some occasions the work would be a selected collection from other works that were titled by their publishers. Listed below are several examples of the first type of extracted works:

- *Night Walk* published by Penguin Classics (2010) — extracted sketch from *The Uncommercial Traveller*;
- *The Hunted Down* by Kessinger Publishing (2010) — extracted from *Hard Times and Other Stories*, vol. XVII;
- *Mrs. Lirriper's Legacy* published by CreateSpace Independent Publishing Platform (2012) and *Holiday Romance* by Dodo Press (2007) — both were extracted from *Christmas Stories from 'Household Words' and 'All the Year Round'* vol. I.

The second type is where different extractions have been collected under a new title not of Dickens's choosing. These types of collections were found to be a challenge when creating the list, as several titles were noted to be already included in other main works.

Examples that demonstrate this issue are:

- *Three Ghost Stories* by Kessinger Publishing (2010), consisting of three items, all of which were taken from *Christmas Stories*. This, for

instance, was one of the documents included in the Giessen University Dickens's Corpus;

- *Complete Ghost Stories* (Wordsworth Classics) published by Wordsworth Editions Ltd (1998), consisting of twenty items extracted from different major works, as detailed below:

Table 4.4 Items extracted from *Complete Ghost Stories*

Item	Description and notes
<i>The Queer Chair, A Madman's Manuscript, The Lawyer and the Ghost, The Goblins who Stole a Sexton and The Ghosts of the Mail</i>	<i>The Pickwick Papers</i> , chapters 14, 11, 21, 29 and 49, respectively), vols. III & IV.
<i>The Baron of Grogzwig</i>	<i>Nicholas Nickleby</i> , chapter 6, vol. VI & VII
<i>A Christmas Carol</i>	
<i>The Haunted Man and the Ghost's Bargain</i>	<i>Christmas Books</i> , vol. XVI
<i>To be Read at Dusk</i>	
<i>A Child Dream of a Star</i>	in <i>Reprinted Pieces</i> vol. XXXIV
<i>The Ghost in the Bride's Chamber</i>	
<i>The Trail for Murder to be Taken with a Grain of Salt and</i>	In <i>The Lazy Tour of Two Idle Apprentices</i> collected in <i>Christmas Stories</i> , vol. I
<i>The Signalman</i>	
<i>Christmas Ghosts</i>	In <i>A Christmas Tree</i> collected in <i>Christmas Stories</i> , vol. I
<i>The Haunted House and Well-Authenticated Rappings</i>	in <i>Miscellaneous Papers</i> , vol. I
<i>The Portrait-Painter's Story</i>	in <i>Miscellaneous Papers</i> , vol. II
<i>Captain Murderer and the Devil's Bargain, Mr Testator's Visitation</i>	<i>The Uncommercial Traveller</i> (Chapters 14 <i>Chambers</i> and 15 <i>Nurse's Stories</i>)

One of the stories here – *Four Ghost Stories* – has since been discovered by Richard Dalby to have actually been penned by Amelia Edwards (Russell 2013), and it was not found in the DCC when searching for some phrases to identify the texts.

Table 4.4 exemplifies how such collections can be misleading in terms of ascribing what appears to be a different work to Dickens, and how some items have been retitled. It is also challenging to discover that some extracted works are interpolated stories in a novel or sketch, where they occasionally have no original title; for example, *The Queer Chair*, *A Madman's Manuscript*, *The Goblins who Stole a Sexton* and *The Ghosts of the Mail* are all from one novel, *The Pickwick Papers*, in chapters 14, 11, 29 and 49, consecutively.

4.4 Identifying the complete works of Dickens

To address this challenge, there are two approaches that have been adopted in order to create a precise and comprehensive list identifying the complete works of Dickens. The precision here refers to identifying the *titles* and their *contents*, to thus avoid duplication in the DCC where some texts could be included on more than one occasion due to differing titles, but identical content.

The first approach that has been adopted features a bottom-up strategy by drawing upon sources that address the works and life of Dickens such as Davis (1999), Jordan (2001), Schlicke (2011), and Shattock (2000). Dickens's works in these references are classified differently, for example by the works' titles being ordered alphabetically in preference to the conventional classification distinguishing between a novel, short story, essay or sketch; occasionally, they are classified as major and minor works. This bottom-up method led to the requirement for inexorable collection of all the details available about each single work.

Consequently, a functional feature of this method is the increased likelihood that precise details can be obtained about each work in question. The references in general offer a brief introduction to the work: when, where and how it was published. Additionally there is a synopsis, commentary,

criticism, and information regarding adaptations; especially for theatrical pieces published during Dickens's lifetime. In some instances, there is indication as to whether the work has been retitled since it first appeared, and where/when it has been published within the collections of Dickens's works.

The second approach to determining the complete works of Dickens is to consider the published collections. This approach adopts a top-down method whereby a collection is first chosen. Then, from the volumes and their contents, along with their available details (chapters, books, sections) as they appear in the published editions, a list is created.

The many editions of Dickens's collected works were published either during or after his lifetime. Shattock (2000: 1181–1273), Schlicke (2011: 207–214) and Davis (1999) provide a detailed survey of Dickens's collections, thus revealing how some editions are 'furnished with introductory or critical matter' (Shattock 2000: 1186), and how they were periodically expended in a chronological manner. Following Shattock's (2000) survey, the collections are of two types: the first refers to the editions published by Chapman and Hall, while the other type refers to the editions of other publishers.

Chapman and Hall, as the then only authorised publisher of Dickens's work in England, published fourteen editions; the first being *The Works of Charles Dickens* (seventeen volumes, 1847–67), with the last being the *Universal Edition* (twenty-two volumes, 1912). The number of volumes in those editions by publishers other than Chapman and Hall comprise twenty-one collections, with several still in progress, such as:

- Penguin English Library edition, later Penguin Classics: (1965-in progress)
- The Clarendon Dickens: Ed J Butt and K Tillotson, Oxford (1966-in progress)

- Norton edition: (1966-in progress)
 - World's Classics edition: (1982-in progress); texts based on the Clarendon edition where available
 - The Everyman Dickens: Ed M Slater (1994-in progress)
- (see Shattock 2000)

As for the purpose of this study, what is essential in these editions is that the core texts have been written by Dickens. According to Shattock's (2000) survey, these editions have no significant differences in terms of the content of Dickens's writings; those which do feature are typically restricted to the technical details of publishing, illustrations, introductions, annotations, character-studies and coloured versions. *The New Oxford Illustrated Dickens* edition in twenty-one volumes excluded Dickens's minor and non-fictional writings with the exception of two items: *The Uncommercial Traveller* and *Reprinted Pieces*. In some editions, the differences include changes to those illustrations that came from the original printing plates and blocks. Others have additional collections of letters or minor works such as *The Life of our Lord*, *Miscellaneous Papers* and other minor items that were not included in other editions (Shattock 2000).

It would be beneficial to obtain and refer to some of the modern editions of Dickens's works, assuming they were completed. According to Jordan (2001), the

best modern editions of Dickens's works are generally those of the Clarendon edition of the novels, published by Oxford University Press ... [that] appear (without textual notes) in the more recent volumes in the Oxford World's Classics editions ... and the Penguin editions are usually the next best, though these are being challenged by a new series of Everymans and ... Norton Criticals, [that] include much supplementary material.

(Jordan 2001: xiv)

However, in her biography of Dickens, Tomalin (2011) stated:

There are innumerable editions of Dickens, Oxford World Classics and Penguin providing some of the best value, many with introductions of high quality. The Oxford Clarendon hardback critical editions of the novels are still far from a complete set and those already published are all out of print, a sad situation.

(Tomalin 2011: 492)

Since all previous published editions are out of print and those modern ones are still in progress, the rational choice was to look for used/out of print editions of the Complete Works of Dickens. Providentially, I found Chapman and Hall's (1906–8) National Edition of The Complete Works of Charles Dickens that I used to determine the total population of Dickens's works.

Based on the National Edition, an accurate list of its entire contents has been created. Although this edition consists of forty volumes, the last two volumes were dedicated to the life of Charles Dickens by his close friend and biographer John Forster. The remaining thirty-eight volumes contain the works of Dickens himself. The contents of each volume have been recorded, adopting the same layout in which they were presented, in terms of chapters, books or parts. The enumeration of volumes was also adopted by dividing the list into thirty-eight main headings representing each volume. *Sketches by Boz*, for instance, is divided into two head titles as it appears in volumes I & II, thus facilitating recognition of how the volumes and their contents appear in the printed edition.

From the two approaches, two lists were produced: the first was created from the references that dealt with Dickens's works and life, while the second was obtained from the printed edition of the *Complete Works of Charles Dickens, National Edition*. A comparison between the two lists was conducted with the intention of creating a final definitive list containing the complete

works of Dickens alongside the contents of each work. This comparison should ensure that both lists supplement each other, where one offers richer and more pertinent details than the other. Post-comparison, the two lists were found to be identical, although additional details were provided. Consequently, a list comprising the sum of both main lists was taken as the comprehensive bibliography of all Dickens's works, based on evidence obtained from the printed edition, in addition to the accurate details regarding each work found in the references and sources that addressed Dickens's works (see Appendix 4.2).

4.5 Dickens's seemingly uncollected works

It should be noted here that while Dickens's works include two volumes, XXXVII and XXXVIII entitled *Letters and Speeches*, these two volumes do not in all respects provide a complete collection of Dickens's letters. Despite the strenuous efforts that have lasted for many decades, and to date have sought to collect Dickens's letters, there is no complete collection as such, and materials continue to surface, albeit with infrequency.

The collections of the Letters of Dickens moved through several stages, extending from the first publication of his letters in 1872–4 until the last collection in 2002; *The Pilgrim Edition of the Letters of Charles Dickens* in 12 volumes. Since 'material still continued to come in', as Easson et al. (2013) indicated, it is relatively unsafe to claim that this list of the complete works of Dickens consists of a comprehensive collection of his letters, at least at the time of writing. The pragmatic reasons for avoiding such a claim are firstly, the impossibility of obtaining an electronic copy of Dickens's letters at the time of writing this thesis. The second reason is related to the copyright, whereby permission is required to store the contents in a retrieval system for

the sake of this study, which was to be discussed upon locating electronic copies.

It also should be noted that there are works that Dickens co-authored or significantly edited with others. Some of these works were included in Dickens's Complete Works, National Edition, such as *Wreck of the Golden Mary*, *The Perils of Certain English Prisoners*, *No Thoroughfare* and *The Lazy Tour of Two Idle Apprentices*, all four of which were co-authored with Wilkie Collins and collected in *Christmas Stories from 'Household Words' and 'All the Year Round'* Vol. I.

These works are to be included in the corpus due to the simple fact that they were included in The Charles Dickens Edition, which was the last edition 'published during his lifetime' where he 'supplied new prefaces and descriptive headlines for the pages' (Davis 1999: 57). Moreover, Nayder (2002) asserted that 'Dickens would take the central authoritative role in the new story, that of the heroic captain, while reserving the roles of passengers and crew members for his subordinates at *Household Words*' (Nayder 2002: 35), where the co-authored works initially appeared. Thus, these limited contributions from other authors were included as they were heavily edited by Dickens's himself, with his style, voice and language apparent enough as to justify their inclusion in this complete works series, and thus in his corpus, the DCC.

On the other hand, 'the fullest and most reliable collections of Dickens's contributions to *Household Word*', as Hagan (1969: 361) suggested, was *Charles Dickens' Uncollected Writings from 'Household Words' (1850–1859)* by Harry Stone. This collection consists of '78 new pieces, either in whole or in part by Dickens, which run to over 600 pages of text and over a quarter of a million words'. Disappointingly, '68 of the 78 pieces are not (as the title

misleadingly announces) 'Dickens' writings', but collaborations – 'composite' pieces, as [Stone] himself called them' (Hagan 1969: 362).

In these sixty-eight pieces, those that can be identified 'as having certainly been written by Dickens himself are only 5, these passages run to only about 4000 words, and their interest is relatively slight'. As for the remaining, 'Dickens's known share in the composite articles to the several pieces – 10, in fact – which external evidence shows he authored entirely by himself, for these are even shorter (about 3400 words)' (Hagan 1969: 363). This results in Dickens's share of the entire collection being only 750 words out of 250,000, that is, only three words in a thousand. Thus, Stone's (1969) collection will not be included in the corpus.

4.6 The construction of the DCC

The DCC is restricted to the texts of Dickens published in the National Edition of his works in thirty-eight volumes out of forty. Therefore, the **only** indispensable criterion for constructing the DCC was solely to have these works fully identified. The rationale for inclusion in the DCC was to precisely create an electronic version of the thirty-eight printed volumes, which obviously required some considered steps and effort so as to match the electronic texts with their printed versions.

One of the essential criteria of a corpus is that it *represents* a particular variety or genre; otherwise it would merely be referred to as a text archive, as differentiated by Leech (1991, cited in Baker 2006: 26) and **Sinclair (2004)**. Being a relatively small and specialised corpus, the DCC represents clearly the original language of the author, and is more likely to contain some oddities in the linguistic norms used by the author, where such curious usage may become lost in the large corpora. Following **a review**, the accuracy of the used to typify the English language of the 19th century, nor the literary

language of that given period in general; rather, it **is representative of** the language of Dickens himself. Meanwhile, the DCC can exemplify the common Victorian era of novels, along with the period in which these works are situated.

4.6.1 Selecting texts' sources

Following the creation of an accurate list of Dickens's works along with their respective contents, gathering the full texts of the complete works has been conducted from one primary source, namely, Project Gutenberg. Other sources were employed as supplementary pools to complete those texts that were not included in Project Gutenberg's electronic versions, such as the University of Adelaide Library, Australia, and the University of Toronto's Robarts Library, both accessed via www.archive.org.

As the comparison progressed between the two lists, the collected texts from these sources were all compared against the printed versions in the Dickens Complete Works, National Edition, in terms of the titles of each work and the table of the contents. This comparison between the electronic texts and the National Edition was to ensure the final corpus contained precise contents with no duplication of either the titles of the works or their contents in general. If, for any reason, the main Project Gutenberg source lacked any electronic texts, these were added manually after first being digitalised and proofread.

The most prominent examples of such absent texts from Project Gutenberg's e-texts database are the introductions that accompany some of the new edition of Dickens's Complete Works, as shown in the final itemised list (see Appendix 4.2). Through completing this task, a precise and accurate comparison has been conducted on two levels: the first comparing the two

lists that enumerate the complete works of Dickens, and the second level comparing the collections of the full e-texts sourced from Project Gutenberg with the National Edition printed versions, in order to prevent any recurrence of the identical texts inside the corpus so as to avoid the multiplication of the number of words either inaccurately or randomly.

The first step towards the construction of the DCC consisted of downloading the complete publically available set of relevant documents from Project Gutenberg to a computer hard-disk. As the available documents related to Dickens are of limited number, the downloading process was carried out manually. Following **a review**, the accuracy of the documents was deemed to be both high and sufficient for inclusion in the corpus. Then, they were reorganised and classified so as to match the National Edition printed version, forming twenty-four files in total. These files were saved to ensure their compatibility with standard corpus tools. Nevertheless, **every effort was made** to match **the contents of these documents** precisely **with** the printed versions of the National Edition of the Complete Works of Charles Dickens, in respect to the thirty-eight of forty volumes concerned with here. The files are classified as follows:

Table 4.5 Contents of the DCC precisely matching the National Edition volumes

File no.	Volume	Work
01	I & II	<i>Sketches By Boz, Illustrative Of Every-Day People</i>
02	III & IV	<i>The Pickwick Papers</i>
03	V	<i>Oliver Twist</i>
04	VI & VII	<i>Nickolas Nickleby</i>
05	VIII & IX	<i>The Old Curiosity Shop</i>
06	X & XI	<i>Barnaby Rudge</i>
07	XII	<i>American Notes For General Circulation and Pictures From Italy</i>
08	XIII	<i>A Child History Of England</i>

09	XIV & XV	<i>Martin Chuzzlewit</i>
10	XVI	<i>Christmas Books</i>
11	XVII	<i>Hard Times and Other Stories</i>
12	XVIII & XIX	<i>Dombey and Son</i>
13	XX & XXI	<i>David Copperfield</i>
14	XXII & XXIII	<i>Bleak House</i>
15	XXIV & XXV	<i>Little Dorrit</i>
16	XXVI & XXVII	<i>Christmas Stories From 'Household Words' And 'All Year Around'</i>
17	XXVIII	<i>A Tale Of Two Cities</i>
18	XXIX	<i>Great Expectations</i>
19	XXX	<i>The Uncommercial Traveller</i>
20	XXXI & XXXII	<i>Our Mutual Friend</i>
21	XXXIII	<i>The Mystery of Edwin Drood – And Master Humphrey's Clock</i>
22	XXXIV	<i>Reprinted Pieces – Sunday under Three Heads</i>
23	XXXV & XXXVI	<i>Miscellaneous Papers – Plays and Poems</i>
24	XXXVII & XXXVIII	<i>Letters and Speeches</i>

Following, where applicable, the principles introduced by Sinclair (2004), the information regarding the texts – i.e. the edition, table of contents, publisher, and introductions by editors – was not included in the plain texts. The text files from the Project Gutenberg eBook customarily contain a description of the file. This pertains to the title of the work, the author, the illustrator, the copyright (that has virtually no restrictions whatsoever on the work's use), the release date, the eBook number, when it was first posted, its language, and finally, the character set encoding. There would also be an indication of the original copy from which the work was transcribed, as well as the proof-reader's details. Yet, with such descriptions and details offering no apparent benefit to Dickens's discourse, these were obviously removed from the text files. The TXT files are intended to represent solely the texts produced by Dickens himself, and any additional details added by editors or publishers who supported the works have been omitted. This describes how

the text files were named, which information was included in each file, and how these text files were stored, that is, the files' format.

4.6.2 Storing and checking

The texts attributed to Dickens in the Project Gutenberg website which were available in TXT format comprised of fifty-five files in total, spanning novels and extracts from various works of Dickens. The TXT format was selected as it is apparent that the majority of text-analysis programmes operate efficiently if the texts are available in such a format, thus ensuring compatibility with the text analysis tools. Each of the main works of Dickens were saved in separate files, as it was suggested by Shattock (2000) that '[i]t is always best to create files at the smallest 'unit', since it is easier to combine files in analysis', while **being** 'stored as individual files rather than as a whole class will allow the most options for analysis' (Shattock 2000: 32–3).

The convention established for naming the files of Dickens's work follows identically the published edition of his complete works, that is, the National Edition as shown in Table 4.5. This resulted in twenty-four files matching the titles and works, as detailed in the DCC's description of contents in Appendix 4.2. This practice thus ensures that the file names relate to their content, the title of the work and the number of the volume in the National Edition.

To the best of my ability, Dickens's texts were preserved as they appeared in the printed version of National Edition, including the non-standard spelling and any ungrammatical structures. These non-standard features of the text may be of interest to some aspects in this study or in futures studies to be conducted on this corpus. The final result was a

compilation of Dickens's Complete Corpus (DCC) which has a total of just over six million tokens (6,202,886).

4.7 Description of the DCC

Unlike general corpora, which are 'normally compiled to be used as a reference for contrastive analysis or to provide a description of the general language' (Rizzo 2010: 3), the DCC is a specialised corpus in the sense that it is limited to texts composed solely by the same author in order to investigate 'its patterns of use' (Biber et al. 1998: 247; see also Xiao & Yue 2009: 241–2 and Litosseliti 2010: 99 for contrasts between general and specialised corpora).

In essence, a specialised corpus refers to a collection of texts (i) designed with specific selection criteria, and (ii) in order to address defined research questions concerned with exploring issues related to the language use where those texts naturally occurred. Other characteristics of specialised corpus include:

- containing 'a particular type [of texts] designed to be representative only of a given type of text, e.g. the Michigan Corpus of Academic Spoken English (MICASE)' (Rayson 2002: 9); The British Academic Written English Corpus, reported by Nesi (2011: 213–228);
- '[featuring] texts that belong to a particular type, e.g. academic prose' (Adolphs 2006: 17). Specialised corpora are primarily 'designed to capture the language of a particular domain rather than the language in general' (Adolphs 2006: 30);
- those 'compiled with the explicit goal of representing a particular variety or subset of a language' (Groom & Littlemore 2011: 153);
- 'collections of texts similar in topic (e.g. medical texts) and type (e.g. academic articles)' (Gavioli 2005: 55);

- it 'includes texts that belong to a particular type, ... the discourse of a particular age group ... the discourse of a particular profession' (Adolphs 2006: 140);
- '[f]ull-text databases of fictional writing can also be used as a basis for principled corpus compilation, resulting in specialised rather than general reference corpora ... which are likely to have a certain idiolectal bias if various works by the same author are included' (Zurich 2008: 180);
- the texts would be 'very similar in their linguistic characteristics (in comparison to other registers)' (Biber et al. 1998: 247).

Sinclair (2004: §2) suggested what he referred to as six 'common criteria' for building a corpus, of which, those most related to the DCC are:

1. 'the mode of the text', since the DCC was primarily recorded as 'written' texts;
2. 'the domain of the text', as a popular and literary text;
3. the 'language' and the 'location' of the texts, where it is British English i.e. the English language variety of the United Kingdom;
4. 'the date of the texts', as being of the Victorian era.

(Sinclair 2004: §2)

4.8 The value of the DCC

The value of the DCC is to provide a means of determining what should be taught to non-native readers in terms of 'high frequency vocabulary items', with this specialised corpus facilitating investigation into the literary language employed by Dickens. Some may argue that learning Dickens's language may not have a useful outcome in terms of language proficiency, as learners of English are not expected to be competent in the English discourse of Dickens's time. Nevertheless, the case is justifiable as understanding the specialised vocabulary of Dickens is necessary to explore specific texts (e.g.

see O’Keeffe et al. 2007: Chap. 10, for the use of high frequency vocabulary items in teaching).

In reviewing the use of corpora for specific teaching purposes, Gavioli (2005) ‘observed that using corpus work with language students seemed to stimulate an ‘investigative’, explorative approach to language learning’ (Gavioli 2005: 53). Baker (2006) suggested that specialised corpora can also be employed ‘to investigate the discursive construction of a particular subject’, and that ‘reference corpora may not contain enough of the text types you are interested in examining or may not have enough references to the subject(s) you want to investigate’ (Baker 2006: 28–9, 31).

Noguchi et al. (2006) highlighted the value of specialised corpus as being to ‘identify features of language that were beyond the scope of thorough observation’ (Noguchi et al. 2006: 156) commonly conducted on general corpus. Identifying these features assists teachers and learners ‘to generate far more objective and reliable data’ for teaching materials, for instance when exploring specific genres. Such data is unlikely to be gathered by merely investigating a general corpus which involves exploring the behaviour of English in general, rather than in a given context, for instance, such as Dickens’s works.

This specialised corpus of Dickensian English also allows investigation of the usages manifested in that context, which further justifies constructing the DCC for the purpose of this study, while exporting the language can facilitate the identification of sources of difficulty for non-native readers of Dickens’s work. An example of such studies that strive to identify a source of difficulty for non-native readers is a study conducted by Umesaki (2000, cited in Noguchi et al. 2006: 156), who found that the ‘variety of referents to the writer in academic papers’ was one source of difficulty for non-native

speakers. Noguchi et al. (2006) also underscored the remaining 'need' for more studies that investigate various disciplinary fields such as 'studying professional English in the theoretical and applied sciences' (Noguchi et al. 2006: 156).

As to why the DCC might be used for teaching Dickens's works, Rizzo (2010) indicated the usefulness of transferring corpora to the classroom where learners 'could directly handle data and learn from authentic samples' (Rizzo 2010: 1). This concept was originally introduced by McEnery and Wilson (2001: 119–121) using 'corpora in the teaching of language and linguistics': 'Such corpora can be used to provide many kinds of domain-specific material for language learning, including quantitative accounts of vocabulary and usage which address the specific needs of students in a particular domain more directly than those taken from more general language corpora' (McEnery & Wilson 2001: 121).

4.9 Conclusion

Following completion of the compilation of the Dickens corpus (DCC), I apply myself to the analyses of the texts, considering various trends that serve learners of English. Concerning the teaching of English vocabulary, many studies have dealt with this from a range of focal points, and it is important to employ those studies so as to serve the purposes of this study. The stylistic dimension in this research is intended to present empirical evidence, consistency and objectivity in terms of the vocabulary selection by implementing stylistic techniques (e.g. corpus stylistics).

Through adopting this method, the impact is thought to be on the study itself, along with any proposed teaching and learning methodology. For example, Gavioli reasons that the analysis of corpora, within the confines of the 'text-types and topics collected', can be a valuable activity to promote

acquisition, especially for learners of English for Specific Purposes (ESP), in order to illuminate those 'features that are recurrent in the corpus itself. Small, specialized corpora ... may provide a source to get an access to uses of specialized language' (Gavioli 2005: 1–2).

I will employ corpus stylistics to introduce readers who may be unfamiliar with Dickensian discourse to some of the most frequent and significant items of the vocabulary of his works. The first stage in the analysis of the DCC will aim to discover which lexical items occur in Dickens's works. The analysis should develop some basic resources such as wordlists or arranging Dickens's novels, for instance according to their vocabulary intensity, rather than their chronological order. Such materials will be produced so as to encourage independent learning intended for learners of English. The wordlists, for example, deal in detail with frequently used words which are common to all of Dickens's works. It will thus be valuable to understand if the analysis of the DCC could reveal oddities in the use of lexical and discourse features. Biber et al. (1998), for instance, suggest that there is no 'general language' as 'each register has its own patterns or use' (Biber et al. 1998: 247). One of the DCC's characteristics may be that it includes different registers for the same author, besides regional and social dialects (see Biber et al. 1998: 247), as considered in the qualitative analysis conducted in Chapter 6.

With the assistance of corpus stylistics' techniques, empirical evidence will be provided to identify the useful 'Dickensian' vocabulary. Consequently, this study will explore the potentiality for improving the pedagogical applications, as well as promoting vocabulary instruction and acquisition by non-native readers. Investigating the construction of Dickens's texts will reveal some patterns, whereby understanding these should facilitate the process of reading and the comprehension of Dickens's works for non-native

readers. Following Gavioli (2005), my 'purpose here is to show that corpus linguistics/[stylistics] theory and methodology involve interesting pedagogical insights which teachers and learners may fruitfully develop in their activities' (Gavioli 2005: 2) to teach the English literature of Dickens.

The dimensions for further investigations include, for instance, organising the DCC's contents chronologically and investigating each piece individually to trace the progress in Dickens's works in terms of lexical density, between say Dickens's first works and later ones, or the development of his style.

Chapter 5 The Applications of Corpus Stylistics

5.1 Introduction

The intention of this chapter is to introduce the essential terms of corpus linguistics that will be applied to the Charles Dickens Complete Corpus (DCC). In doing so, the concepts of *wordlist*, *keyword*, *collocation*, *cluster* and *semantic prosody* will be introduced and defined. This will enable the application and interpretation of these concepts to be reported in Chapter 6, in addition to the pedagogical extensions where these can be suggested, with a particular emphasis placed on the non-native readers of Dickens's works.

Highlighting three major methodological approaches – keyword analysis, identifying typical extended lexical phrases, and collocational analysis – Biber (2011) confirms that the majority of the studies on corpus stylistics place the lens of focus on how words are distributed, in order to locate those textual features that can be deemed to be indicative of a particular author, a text, or a character placed within a novel or play. Moreover, Fischer-Starcke (2010) asserts that language is by its very nature an open system and that within a corpus there can only be found a limited selection of the potential language varieties being considered. Therefore, any language-based assertions tend to generalise those features that can be found in the corpus, while revealing the probability of such features occurring in the language varieties considered by the corpus. In the case of the DCC, which includes all the works of Dickens based on the National Edition of his complete works, there would be scope to make 'absolute statements' regarding Dickens's language contained within the DCC, which consists of all his novels, novellas, short stories and letters, as detailed in Chapter 4 (e.g. see Fischer-Starcke 2010: 14).

Notwithstanding considerations of a theoretical nature, firm and pragmatic grounds exist for the attraction of research through word-based corpora, regardless of **the** grammatical construction. This relates to annotation. Study grounded in words can be conducted with raw corpora by employing software programmes that identify word forms and provide examples in use in concordance (cf. Kennedy 1998: 8); however, limited grammatical data may well restrict the research breadth and effectiveness. In order to conduct research based on categories, corpora should ideally feature annotations for syntactic features and grammatical structures. These may well require significant periods of time to develop, and so can be feasibly substituted by corpora annotations that focus on the grammatical characteristics of words such as the parts of speech. Nevertheless, even in a corpus that is grammatically annotated, the ability to deduce concordance from a verb such as *play* in all its respective forms will be easier than gaining a concordance of every past perfect construction, for instance. In cases when employing a raw (i.e. untagged) corpus, resolving the former would be somewhat time-consuming, while resolving the latter would demand a more significant investment of time.

Halliday (1992, cited in McEnery & Gabrielatos 2006: 35–40) presents a summary of the pragmatic considerations regarding research reliant on words and categories, and in respect to the ease of observing and separating the data employed by lexicologists due to their form as words of lexical items, as opposed to the data used by grammarians, which are more challenging to access and severely tested the software potential of the era.

In the following sections I outline the primary analytical techniques that I employ in my analysis of the DCC.

5.2 Wordlists

Baker, Hardie and McEnery (2006) state that a wordlist is '[a] list of all of the words that appear in a text or **corpus**, [and is] often useful for dictionary creation' (Baker, Hardie & McEnery 2006: 169; bold in original). They **state** that wordlists typically provide data on how frequently a word or token can be found in a corpus, with such derived lists usually being presented in alphabetical order or in respect to their frequency in terms of total aggregated frequency or the proportion that the words represent in respect to the entire text. Furthermore, 'word lists can be **lemmatised** or **annotated** with part-of-speech or semantic information (including probabilities – for example, the word *house* occurs as a noun about 99 per cent of the time and as a verb 1 per cent of the time) (Baker, Hardie & McEnery 2006: 169; bold in original), and are necessary when evaluating the number or importance of keywords. McEnery and Hardie (2012) describe that 'a word frequency list [is one] which lists all words appearing in a corpus and specifies for each word how many times it occurs in that corpus' (McEnery & Hardie 2012: 2), while Hunt and Carter (2012) assert that the primary function of a wordlist is to calculate the raw frequency of word types found in a corpus alongside the aggregate number of running words, and that by appraising this data, such as the most frequently occurring verbs or nouns, it is possible to begin to develop some understanding of the prominent linguistic characteristics of a text. Hunter and Carter (2012) move on to highlight the potential to enhance the utility of such raw frequency data through employing a tool referred to as KeyWords, which establishes in an objective manner those words which appear in a corpus with high frequency in comparison to significantly larger reference corpora. The analysis of keywords thus illuminates those words which establish the 'aboutness' of the text, along with recurring names and concepts that stand a text apart from its peers and through which the text's meaning can be formed (Hunt & Carter 2012: 30).

Scott and Tribble (2006) point out that a wordlist is essentially a list that comprises of various word types. The wordlist programme scans through an entire text and forms types from all the repeated tokens, 'that is, each instance (*token*) of the word **the** is counted but the completed list displays the only once as a *type*, usually together with its frequency (the number of tokens found)' (Scott & Tribble 2006: 12–3; emphasis in original). To extend our understanding of the WordList tool further, Hunt and Carter (2012) explain that during its primary pass through a text, the tool rapidly aggregates the total instances of each word. In their case they used *The Bell Jar* as the sample text, and underscore that grammatical words such as *the* and *a/an*, as well as *and* and *to*, were the most frequently occurring words in the book. Hunt and Carter (2012) mention that such function words permeate through most English texts and in themselves offer little value in semantic meaning. Semantically, there is an important contrast between content words and function words. Content words that include nouns, lexical verbs, adjectives, and adverbs have information and meanings e.g. nouns indicate objects and verbs indicate to actions happening. Regarding the other function words such as propositions, determiners and pronouns, despite these having a major influence grammatically, their lexical meaning is either limited or ambiguous, merely serving to indicate the structural relationships that words have between them (see Graesser et al. 2004: 197). Nevertheless, by passing down through the list they identified frequently occurring content words which offered the potential to unveil recurring patterns in the protagonist's narrative and thus provided insights into her plight. The resulting wordlist could also be utilised to create a comprehensive table of the keywords featuring in the novel, and then allow the software to make a comparison between those and a reference corpus, once the proper nouns of characters and places had been removed.

It has been indicated by Scott and Tribble (2006) that while wordlists initially feature a limited number of high frequency items, they then progress down towards a huge number of words which may only appear with rarity in a corpus. One of three issues associated with keyword lists as discussed by Scott and Tribble (2006) relates to the 'transformational' feature, whereby an individual or a group of texts is recast into a different form, shifting the object of consideration fundamentally 'from a text which can be read linearly to some other form which will give rise to important insights, pattern recognitions or teaching implications' (Scott & Tribble 2006: 12). Such a shift leads the reader to focus on the form or nature of the words, as opposed to the overarching message conveyed through the text.

Regarding the potential of wordlists, Bowker and Pearson (2002) point out that by employing the online application *WordLister* to scrutinise lists based on words' frequency and alphabetical order, it is possible to begin to acclimatise to the type of language contained within a corpus. Moreover, frequency and keyword lists can form a foundation for an initial appraisal of a text or corpus, highlighting those lexical items that feature with frequency, and revealing something of the meaning embedded within the text. For example, through identifying a word in a list, the user may be motivated to develop a concordance to discover its use in greater depth. There are additional benefits to clusters of words that enable the identification of larger units, as through considering the latter, similarities may become apparent that stimulate further investigation. Similarly, it may become clear that particular verbs 'appear to co-occur more frequently with one preposition rather than another. Again, you will need to use the concordancing tool in order to investigate further' (Bowker & Pearson 2002: 119).

Therefore, wordlists offer a range of applications, such as becoming familiar with a corpus's terminology, understanding how particular words

signpost an interesting aspect, but rather that the important points to be considered can be found in the relationship between the style and keywords. Moreover, Culpeper (2009) asserts that the notion of the keyword is well established through enjoying a more than 50 year history and more recently the infancy of keyword analyses in the 1990s, with the key innovation in recent times being the development of software programmes that can conduct the required analyses effortlessly. As a result of such developments, the modern researcher can swiftly and relatively easily 'calculate the incidences of each and every single word in the target data as well as a comparative data set, undertake statistical comparisons between incidences of the same words in order to establish significant differences' and access the final keywords ordered in respect to their 'degrees of significance of difference' (Culpeper 2009: 30).

In establishing space between language and text, mind and culture, Scott and Tribble (2006) focus on study that is centred on language, so while a word may enable association and meaning within a text, that does not necessarily suggest that the word itself has particular significance in the wider English language. Culpeper (2009) defines the keyword as 'simply a term for statistically significant lexical items' (Culpeper 2009: 32), while identifying with certainty the relationship between keywords as markers of style and sequences in particular texts through the assertion that style markers as words that differ in frequency considerably when compared with their normal usage accurately reflects the nature of keywords. Furthermore, it is asserted that while repetition underscores style markers and keywords, it is 'only repetition that statistically deviates from the pattern formed by that item in another context' (Culpeper 2009: 33). To identify keywords thus requires consideration of repetition, with the notion being that those words that re-

occur with frequency in texts are most likely to be fundamental to such texts (Scott & Tribble 2006).

Baker et al. (2006) caution that keyword analysis may lead to the erroneous belief that words appearing with frequency are somehow inevitably essential to comprehending the text, by virtue of their prominence in a number of texts within a corpus. They offer the following example: 'a corpus consists of 1,000 equal-sized files, and the word *ironmonger* only appears fifty-eight times in one single file called 'The history of the ironmonger', whereby a keyword analysis may reveal this word to be considered as important across the texts (Baker et al. 2006: 97). Consequently, in establishing the relevance and importance of a word, a distributed search through an entire corpus or lists of keywords should be considered, to compensate for the obfuscating nature of 'disproportionate representation'. The keywords found in texts reflect striking repetition due to their significant use as word forms when statistically compared with their incidence in a comparable reference corpus (Toolan 2008). It is presumed that where keywords do feature, as this is not a certainty in texts, they will act as a signpost 'to a text's themes or preoccupations, and that by virtue of their recurrence they cannot easily be ignored by the reader' (Toolan 2008: 12).

Regarding the notion of being *key*, Scott and Tribble (2006) underscore that it is common for languages to feature the word metaphorically to highlight the importance of individuals, locations, words and thoughts; with the term being employed in textual databases to retrieve, and that in essence it does not require definition. The nature of "'keyness'" is, therefore, a quality which is generally intuitively obvious' (Scott & Tribble 2006: 55–6), although it can be said to reflect the value that those words may have been assigned in a text (implying importance) and being highly relevant to the core of the text, while circumnavigating unnecessary detail. Stubbs (2010) asserts that

'keywords are words which are significantly more frequent in a sample of text than would be expected, given their frequency in a large general reference corpus' (Stubbs 2010: 25–6), establishing that (i) keyness is concerned with texts, as well as text-types and intellectual domains; (ii) software converts texts into wordlists or 'lists of n-grams' before engaging in comparison of these lists with other reference corpora, and then through the organising and filtering of such lists, simple patterns can be found despite the volume of text; and (iii) content words have a strong correlation with the 'propositional content of texts' (Stubbs 2010: 25–6), and despite keyness being associated with texts, the nature of the list-formation process means that resulting patterns do not concern the segmentation of texts, being a facet of 'global textual cohesion' (Stubbs 2010: 25–6), rather than the structure of texts.

Those keyword lists that result from corpora searches feature 'two types of key word: "positive" (those which are unusually frequent in the target corpus in comparison with the reference corpus), and "negative" (those which are unusually infrequent in the target corpus)', with these terms corresponding to 'overuse' and 'underuse' in the corpus literature for learners, for example (Rayson 2008: 523). It is important to remain mindful of the alternative senses for keywords: the **cultural sense**, as being reflective of a focus for the organisation around cultural fields (Wierzbicka 1997, cited in Stubbs 2010: 23); the **statistical sense**; and the **figurative sense** (Stubbs 2010) (see also Section 3.1.3). O'Halloran (2007) claims for the benefits of keywords in enabling meaningful and pertinent understanding of texts, while reminding that such keywords may not be those that feature with the greatest frequency, as the definite article, for example, may appear with the greatest frequency but will not offer any insight into the text, and thus can be rejected as a keyword. **This justifies the exclusion of function words when analysing the DCC's keywords.** Meanwhile, Scott (1997b, cited in O'Halloran 2007: 233)

argues that keywords offer a valuable opportunity to begin engaging with a text. With keywords being the product of an undistorted and unbiased process, the technique used to extract them can be confidently said to be objective (O'Halloran 2007).

The software programmes employed to produce keyword lists systematically compare frequency patterns between a pair of texts of corpora, with keywords being deemed to be positive when featuring with greater frequency than would be expected in relation to the reference corpus, and deemed to be negative when featuring with less frequency (Bondi 2001). Although it is possible to gather a significant body of information from wordlists (particularly when collocational and cluster details are included) such as the frequency and patterns of occurrence, frequency wordlists fail to inform on words that appear with atypical frequency or infrequency; therefore, the analysis of keywords through applications such as WordSmith Tools facilitates the exploration of such words (Tribble 2001).

A wordlist can also be said to be an indicator of what is important and requires greater attention in terms of teaching (**see Table 6.14**), for instance, and understanding what a text is *about*. The 'aboutness' of the DCC will be addressed in the analysis found in Chapter 6. Mahlberg (2010) believes in the value of keywords and their ability to offer an introduction to a text, 'pointing to words that are potentially useful for more detailed analysis' (Mahlberg 2010: 296), while Bowker and Pearson (2002: 114–5) report that although this is the case, corpus tools such as WordSmith Tools unlock the potential to locate words appearing with an atypically high frequency in a text or corpus when compared with a different corpus; this being referred to as 'key types'. Therefore, the advantage of keywords over frequency lists is this ability to make comparisons with other selected corpora or texts in order to identify words that can be considered to be highly relevant. Essentially, the keywords

are graded due to their keyness rather than frequency, so that they can be highlighted for further consideration (Bowker & Pearson 2002). The analysis of keywords offers the potential to illuminate the general features and traits of the language employed by Dickens.

5.4 Collocations

5.4.1 Defining collocation

Nesselhauf (2004) explores the notion of collocations in order to offer ‘a systematic overview of the widely varying definitions of the term as well as of a number of related terms such as “selectional restrictions” and ... an overview of the ways in which collocations have been classified’ (Nesselhauf 2004: 11–34). She then presents her definition and classification of collocations, which attempts to ensure the development of language data through both theoretical consistency and easy application, **employing the phraseological definition rather than the frequency-based one**. Seretan (2011) defines collocations by considering both the statistical and linguistic approaches, finding that collocation as a term has traditionally been employed with a wide scope to generally describe words that co-occur. This initial statistical perspective is then contrasted against a more linguistically-centred perspective in positing that collocation items are related syntactically. With this latter perspective gaining increased agreement, ‘some authors have suggested to use distinct terms to distinguish between the two understandings’ (Seretan 2011: 15).

Given the diverse employment of the term collocation, Nesselhauf (2004) identifies two main spheres. In the first, the approach considers a collocation to be words that co-occur with a particular distance, with discrimination made where the co-occurrences are frequent. The second

stance is that collocations are considered to be a form of word combination, and typically one that has certain rigidity though not being comprehensive. The first stance, also referred to as the 'frequency-based approach', extends back to Firth, having been further developed by Halliday and Sinclair, among others, who are concerned with the relations of syntagmata through computational research; while the second stance, also referred to as the 'phraseological approach', has been significantly shaped by Russian phraseology, with those researchers engaging with this second approach including AP Cowie, I Mel'čuk and FJ Hausmann (Nesselhauf 2004: 11–2). For the purposes of identifying the collocations in the DCC, the approach adopted for this study is the frequency-based approach (see Jeffries 2006).

Hoey (2005) highlights that the notion of collocation is typically credited to Firth, and that his discussions form the foundations for everything that has subsequently contributed to the topic.

Interestingly, though, Doyle (2003) draws attention to the fact that the word *collocation* was being used in linguistic discourse prior to Firth; in this connection he draws attention to a citation from 1940 in the *Oxford English Dictionary* (1995). This observation is confirmed by inspection of the 1928 edition of *Webster's New International Dictionary*, which has the following entry for *collocation*: Act of placing, esp. with something else; state of being placed with something else; disposition in place; arrangement. The choice and *collocation* of words. ... Collocation denotes an arrangement or ordering of objects (esp. words) with reference to each other.

(Hoey 2005: 3; emphasis in original)

Therefore, collocation can possibly be referred to as a characteristic of language through which two or more words co-occur with frequency (e.g. *exceptional + circumstances*) (Hoey 2005). More in-depth discussion on the term is available in Hoey (2005) and Sinclair (1991: Chap. 8), with the latter defining that collocation refers to the incidence of two or more words in close

proximity in a text. Baker et al. (2006) describe collocation as the increased likelihood that some words will co-occur with other words in particular contexts, and a collocate as being a word that can be found in the vicinity of another word.

Collocation reflects the relationship found within a syntactic unit and among those distinct lexical components that typically co-occur in certain scenarios and feature complete transparency from a semantic perspective (Sterkenburg 2003). Moreover, collocation is frequently described as a defining word association 'whose lexical constituents have developed an idiomatic relation based on their frequent co-occurrence' (Burkhanov 2003: 109). The differences between collocations and lexical bundles, or clusters as termed by Scott (2011), is that lexical bundles can be 'words which follow each other more frequently than expected ... "extended collocations" and contain grammatical (functional) words, e.g. *as a result of* or *in pursuance of* (see Hyland 2008) while collocations refer to 'the relationship that a lexical item has with items that appear with greater than random probability in its (textual) context' (Hoey 1991: 6–7, cited in Hoey 2005, emphasis in original). Therefore, collocations '[t]ypically ... are expressions that can be interpreted more or less correctly out of context, but cannot be produced correctly if the conventional expression is not already known to the speech community' (Croft & Cruse 2004: 250).

Boers (2009) refers to the popular definition from Alison Wray (2002) for collocation as 'formulaic sequences' of 'lexical phrases' or 'chunks' (Boers 2009: 3) that appear in a sequential order that may be continuous, of words, phrases or other elements that seem to be pre-constructed, and are thus cognitively stored and retrieved from the memory at the point of use, as opposed to being subject to creation and analysis through the grammar of the language. Regarding the chunks, the *Longman Dictionary of Language Teaching*

and Applied Linguistics defines these as 'a unit of language that forms a syntactic or semantic unit' while additionally featuring an 'internal structure', such as (i) a stream of text that extends beyond the length of a sentence but falls short of the expected length of a paragraph; and also (ii) a stream of text that is shorter than a sentence and performs a comprehension and production-related function (Richards 2010: 76–7).

Regarding lexical collocation, it has been claimed that this is determined by the usage or those syntagmatic relationships that are deemed to be preferable between two lexemes in a syntactic pattern (Granger & Paquot 2008: 43), whereby both lexemes contribute semantically to the co-occurrence but have different lexical statuses, with the 'base' of the collocation being initially chosen due to its denotation, and the second collocating component being chosen due to its semantically dependent relationship with the base. **Such observation is related to Firth's (1951) famous assertion that 'you shall judge a word by the company it keeps' (cited in Hyland 2008: 5).** Moreover, Renouf and Sinclair (1991) present the notion of collocational frameworks as being non-continuous two word sequences located adjacently, that cannot be considered to be grammatically independent as the integrity of their structure is reliant on what then intervenes.

Gardner (2007) conducted a seminal survey on the effect of multi-word items on the lexicon of the English language, with a major corpus linguistic-related finding being the significant role of lexical collocation, while he also found that when considered broadly, collocation features two or more words which feature together with strong frequency in the confines of a limited area of discourse. Particularly pertinent to this theme are the examples of collocation that are well established within a language, that are

often referred to as formulaic language (Wray 2002), formulaic sequences (Schmitt 2004), or multi-word items (Moon 1997) ... : "A multi-word item is a vocabulary item which consists of a sequence of two or more words (a word being simply an orthographic unit). This sequence of words semantically and/or syntactically forms a meaningful and inseparable unit' (Moon 1997: 43).

(Gardner 2007: 255)

The key characteristic of multi-word items is that they are comprised of additional word forms, so that in their entirety they encompass a significant proportion of the English language in both its spoken and written forms. Moreover, it has been estimated that multi-word items comprise over half of all spoken and written English, which it is suggested leads to the status of such idioms and multi-word collocations as being linguistically significant (Erman & Warren 2000, cited in Gardner 2007: 255).

Evidence from different trajectories also implies that the volume of multi-words items could be greater than the sum of all the words in the English lexicon, while numerous multi-word items have two or more definitions, as exemplified in the British National Corpus where it is asserted that the one hundred most utilised phrasal verbs have over five hundred potential definitions, thus casting doubt on the validity of corpus-based research that is essentially dependent on individual word form frequencies (Gardner 2007: 255).

For Sinclair (2004c, cited in Gardner 2007: 255) and his colleagues, a less conventional perspective to multi-word items is adopted, indicating that the majority of lexical meaning is more closely related to the patterns of words than their individual states. This has been termed the 'maximal approach', which allows the scope of units of meaning to embrace all the relevant associated words that are deemed to have a relationship and connection to the central word, as Sinclair asserts that a lexical item comprises

of all those words that collaborate to form the unit of meaning (Sinclair 2004c, cited in Gardner 2007: 255).

From a theoretical perspective Geeraerts (2010) posits that collocation, defined as the lexical relationship between a pair or more words in which their co-occurrence in texts is in close proximity, may adopt a range of forms inclusive of *collocation*, *colligation*, *semantic preference* and *semantic prosody*.

In **collocation** the base target word is typically referred to as the node, with the adjacent word that occurs with frequency being referred to as the collocates (Geeraerts 2010). The node that is subjected to collocational analysis could potentially be a word form or word in the case of the application of lemmatisation, and the nodes may also be in the form of a phrase or expression (Geeraerts 2010).

In respect to *colligation*, Smith (1996) describes this as grammatical selections co-occurring with frequency, that is, the type of syntactic pattern manifesting due to a word's inclusion, such that 'co-occurrences, in other words, are now defined between the node and a syntactic class' (Smith 1996, cited in Geeraerts 2010: 170). **Therefore, colligation refers to the relationship between words at the grammatical level (see Xiao 2009: 996).**

Semantic preference is defined by Geeraerts (2010) as being placed within a medium level of abstraction and within the boundaries of 'syntactic colligation and lexical collocation', while being concerned with the relationship that exists 'between the node and a set of semantically related words' (Geeraerts 2010: 172). Semantic prosody considers the nature of co-occurrence not from the lexical or semantic standpoints, as in the cases of collocation and colligation, respectively, but rather from the perspective of connotation, that is, from the emotional or evaluative voice of the adjacent lexical items, while referring 'to the fact that words may have a tendency to

line up with either positively or negatively evaluated words’ (Geeraerts 2010: 172).

5.4.2 The significance of collocation

Stubbs (2001: 24) defines collocation as ‘a lexical relation between two or more words which have a tendency to co-occur within a few words of each other in running text’, confirming that his definition therefore represents ‘a statistical one: “collocation” is frequent co-occurrence’ (Stubbs 2001: 29). For example, in the DCC, the node collocate *write* was found to collocate, for instance, in terms of time with *morning* more than with *night* (with the minimum frequency set to 10 times) (see Table 5.2 below).

Table 5.2 The collocates for the word ‘write’ in the DCC (using AntConc 3.4.4w)

No.	Rank	Freq	Freq (L)	Freq (R)	Stat	Collocate
1	102	14	7	7	4.77502	morning
2	114	13	5	8	3.51864	day
3	130	11	5	6	3.38137	night

Investigating the concordances lines for the three collocates can reveal further details regarding the interpretation of such collocates, representing an attempt to respond to questions such as: Why does the node collocate ‘write’ appear with *morning* in greater frequency than with *night*? Does this reflect a custom for the writer himself, that is, Dickens, or rather a custom for a character in his work? Is this perhaps related to a preference for writing during daylight hours? Might this be reflective of the historical period where electricity was not yet used domestically in England?

In corpus linguistics the concept of ‘collocation’ may be used as a relatively neutral descriptive tool to focus on surface features of language. The term ‘collocation’ was already in use before the advent of electronic corpora to describe words in their contexts. (The term ‘collocation’ is typically associated with the work of J. R. Firth).

(Mahlberg 2005: 22)

Wray's (2002) description of **collocation** encompasses the full range of word strings that can be presumed to be considered as whole units by users of any particular language, thus allowing for the real possibility that certain such word strings seen to be conventional for one language user may not necessarily be so formulaic to other language users (Boers 2009). Essentially, the range of chunks that a particular interlocutor may have within their linguistic reach may well be particular to that individual and thus vary between others, which Boers (2009) explains to be as a result of the accumulation of such repeating chunks of language by native speakers from childhood onwards as they receptively encounter them, and that this phenomenon of exposure to word strings is entirely unique for each speaker. Such variations in exposure may be as a result of the physical location and language variety spoken in that region; the shifts in language use as it develops over time; disparate interests leading to differing schemata; and distinct discursal encounters due to academic, professional or vocational influences, etc. (Boers 2009). Consequently, such 'word strings that are experienced as formulaic (and idiomatic) by members of a particular group of language users' and have the potential to define them as members of a particular group 'will not always be perceived (or mentally processed or remembered) as such by people outside that group' (Boers 2009: 4).

Hill (2002) emphasises the importance of collocations from a pedagogical point of view, while stating that (i) rather than lexicon being arbitrary, vocabulary can be predicted to an extent; (ii) collocations can form patterns that can assist in learning through the phenomenon of predictability; (iii) there is the potential for as much as seventy per cent of all text that occurs naturally to be located with a fixed expression of some form or other; (iv) memory serves to allow the retrieval of collocates from our mental lexicon; (v) collocation enhances fluency through enabling the ability to think and

communicate with efficiency; (vi) complex ideas can be expressed with greater ease through complex noun phrases that comprise of relatively common words; (vii) collocation eases thought processes by allowing the speaker to convey complex ideas quickly, thus freeing up mental space for adapting ideas; (viii) the adept use of stress patterns for known collocational phrases will enhance the speaker's use of intonation and stress; and (ix) that the recognition of chunks of language is essential for learning, with learners enjoying enhanced acquisition if they correctly hear chunked text (Hill 2000).

5.4.3 Exploring collocation through corpus methodology

Cortes (2002) discusses that for a number of decades, research into the patterns of lexical association have aroused interest in linguists, namely the manner in which words appear to feature in close proximity and association with other words. Furthermore, Firth (1964, cited in Cortes 2002: 131) coined the terms 'collocation' and 'collocability' to enable the expression of the common co-occurrence of words, and with the intention of describing the manner in which collocates determine the final meaning of words.

The steady increase in the study of corpus linguistics and the availability of computer software specifically dedicated to analysing language corpora through the introduction of innovative research instruments and directions of interest contrasts sharply with the previous empirical landscape, where research into the frequency of lexical co-occurrences was more qualitative and grounded in perception rather than hard data (Cortes 2002). Research methodologies have been utilised to explore the spread of patterns of lexical association in raw environments, enabling both the identification of the arising collocation and the aforementioned lexical bundles that feature lengthy collocational patterns of up to six words that can be found to co-occur

with statistical significance in register (Biber et al. 1999, cited in Cortes 2002: 131).

The concept of collocation as considered by Flowerdew (2009) relates to whether words usually co-occur or not, and if so, how they tend to co-occur, and while software programmes can provide the user with a list of collocates, those patterns that occur with frequency will highlight common collocations. An example of this is offered by Hunston (2002, cited in Flowerdew 2009: 113), whereby '*shed* ... collocates with *light, tears, garden, jobs, blood, cents, image, pounds, staff, skin, and clothes*. Typically different collocates will affect the precise meaning of the word, e.g., *shed blood* means to suffer, *shed pounds* means to lose weight, and *shed image* means a deliberate changing of how one is perceived' (Hunston 2002, cited in Flowerdew 2009: 113). **This concept is related to the attitudinal meaning investigated in semantic prosody, e.g. the verb *cause*, which is used negatively with, for instance, the nouns *accident* and *catastrophe*.**

Collocations are often explored through the creation of a text or group of texts' concordance, with a common approach to mapping concordance being by way of the 'Key Word in Context index (or KWIC-index)', which can offer insightful illumination on the manner in which words have been employed in the target source (Geeraerts 2010: 170).

Baker et al. (2006) advise that the WordSmith Tools application enables users to define a specific window within which to calculate the frequency of collocations, and they provide an example table using the Brown Corpus, where the first ten collocates for a particular word are queried to 'within a -5 to +5 span' (Baker et al. 2006: 37), though Sinclair talks about a 9 word window.

The techniques available for use in corpus linguistics have thus enabled researchers to prove both the frequency and likelihood of certain collocates, through employing ‘statistical methods such as **mutual information**, the **Z-score** (Berry-Rogghe 1973), **MI3** (Oakes 1998: 171-2), **log-log** (Kilgarriff and Tugwell 2001) or **log-likelihood** (Dunning 1993) scores’ (Baker, Hardie & McEnery 2006: 37; bold in original). Utilising each method will result in a returned value that indicates the collocational strength, although the criteria applied will vary. The mutual information method, for instance, returns the frequency that collocates will co-occur in contrast to occurring individually, while log-likelihood will report a strong collocation where the independent words appear with frequency. Therefore, mutual information queries will return ‘a high collocation score to relatively low-frequency word pairs like *bits/bobs*’, while log-likelihood queries will return ‘a higher score to higher frequency pairs such as *school/teacher*’ (Baker, Hardie & McEnery 2006: 38).

By embracing the capability of modern computers, it is now possible to measure co-occurrence and define collocation statistically, namely that collocation is deemed to be so if the two words co-occur with greater frequency than would be reasonably expected in the natural distribution of the individual words (Mahlberg 2005). A variety of statistical approaches are employed to establish collocation that is significant, with the tools available through corpora enabling the querying of collocations as a standard feature (Mahlberg 2005), e.g. *a case of extreme destitution, an object of sympathy or an increase in popularity*.

With the collocational focus often being placed on lexicon, Renouf and Sinclair (1991) notice that grammatical words are often the most common candidates for collocation in a language, while they demonstrate this by describing ‘frameworks which consist of a discontinuous sequence of two

words, positioned' at a location one word removed from one another, such as 'a + ? + of' and 'an + ? + of' (Renouf & Sinclair 1991, cited in Mahlberg 2005: 22).

An alternative approach to the consideration of words found in a sequence that co-occurs with frequency can be referred to as the 'lexical bundle', typically defined quantitatively and more often than not failing to comprise of a structural unit, with Biber et al. (1999, cited in Mahlberg 2005: 22) being primarily concerned with how these lexical bundles are distributed between different registers. The notion of collocation therefore models clearly how software applications and their descriptive functions have the ability to describe the various theoretical positions in corpus linguistics (Mahlberg 2005). Consequently, it can be said that software tools and their descriptive functions have the ability to describe the various theoretical positions available in corpus linguistics concerning the concept of collocation. These positions consider 'the distribution of lexical bundles across registers' or 'interpret[ing] the elements that are found in the middle of a frame functionally, and show that frameworks are highly selective of the words that complete the sequence' (Mahlberg 2005).

5.4.4. The application of collocation in language teaching

Collocation offers useful potential in respect to the teaching of languages, through awareness-raising of collocates that appear with low frequency and which have been archived by native speakers, while collocates can also facilitate awareness-raising regarding the bias or connotation that may exist in words (Baker, Hardie & McEnery 2006). Nesselhauf (2003) confirms that while collocations are of significant value for those learners targeting 'a high degree of competence in the second language', they are also of merit for those learners with more pragmatic aspirations, since they enhance both accuracy and fluency (Nesselhauf 2003: 223). This can be

demonstrated by considering the strongest collocate of the word *bystander* in the British National Corpus, which is *innocent*, indicating ‘that even in cases where *bystander* occurs without this collocate, the concept of innocence could still be implied’ (Baker, Hardie & McEnery 2006: 38). In this regard, Karoly (2005) asserts that to know a word is undoubtedly more than merely understanding its definition, with Nation (1990) proposing eight factors that would need to be fulfilled in order to ensure comprehensive knowledge of the word: (i) spoken form, (ii) written form, (iii) grammatical behaviour, (iv) collocational behaviour, (v) frequency, (vi) stylistic register constraints, (vii) conceptual meaning, and (viii) word associations (Nation 1990: 31).

Flowerdew (2009) continues by posing the question of how best to respond to a language user who incorrectly collocates, suggesting that an explicit correction can be made, where the learner is informed using grammatical nomenclature of the reason why the particular collocation was not possible. The language repair is suggested to be more effective and profound; however, the learner is encouraged to explore the concordances of the word in order to accumulate a selection of phrases, and then to explore those phrases to discover if there are inherent clues in the collocations that are valid and that will help to inform the learner in future, with this representing a more valuable and memorable learning experience. Finally, Flowerdew (2009) points out that the software used to construct concordance can support educators, particularly of lower proficiency non-native learners, where reliance on intuitive understanding is less realistic.

5.5 Lexical bundles, formulaic expressions and key clusters: interrelated phenomena

5.5.1 Lexical bundles

Biber (2006: 133) **underscores** that the expression 'lexical bundle' was first coined in the *Longman Grammar of Spoken and Written English*, which conducted a comparison of recurring patterns of words and phrases used conversationally and in academic parlance, with the framework then being employed in a number of later research projects.

Biber et al. (1999, cited in Cortes 2002: 134–5) consider lexical bundles as expressions that recur regardless of their 'idiomaticity' and structure, and as representing basic word sequences that can be found to co-occur when language is used naturally, and that while 'three-word bundles' should be viewed as collocations, four to six word bundles should be considered as having a more phrasal rather than collocational nature, and occurring with less frequency – which is an important characteristic for lexical bundles. Biber et al. (1999, cited in Cortes 2002: 134–5) limit the bundles they examine to those occurring with a frequency of ten instances per million words, and being found in five or more texts, while they assert that the primary difference between these lexical bundles and other lexical association patterns is the nature of the search process, which ignores established perceptions and intuition and relies solely on the prescribed search criteria.

In order to identify lexical bundles the strategy must be grounded in frequency, as lexical bundles are essentially those word sequences that appear with the most frequency. Biber (2006) offers the two examples of '*do you want to*' and '*I don't know what*' as being typically representative of those features found in lexical bundles, where it can be seen that (i) the meaning is literal as opposed to being idiomatic, and (ii) the structures are not typically complete

from a grammatical perspective (Biber 2006: 133–5, *italics in original*). With the established frequency count to define patterns of words as being classifiable as lexical bundles being subjective, in Biber's (2006) analysis the threshold is set somewhat cautiously at 40 instances for every million words analysed.

Biber's (2006) analysis navigates through the issue that a great number of the identified lexical bundles feature with considerably **higher** frequency than the conservatively set threshold, at perhaps over two hundred instances for every million words analysed, and so to pragmatically define the breadth of the study he restricts the number of sequences to be identified to four word lexical bundles.

An additional feature of lexical bundles that can be considered to be defining is the requirement for the sequence of words to feature in five or more individual texts in order to be considered as a lexical bundle (Biber 2006), as this requisite moderates the stylistic use of multi-word sequences by particular authors or speakers, and acknowledges that the majority of lexical bundles can be found dispersed in the texts included in a corpus, while those most atypical lexical bundles were typically distributed in twenty or more texts in Biber's (2006) analysis of classroom tuition and coursebooks.

In respect to the meaning embedded in lexical bundles being non-idiomatic, the meaning of two aforementioned sample multi-word patterns (*do you want to* and *I don't know what*) can be considered as being clear from the singular words the bundles contain, although Biber (2006: 134) reveals 'that bundles typically function as a unit in discourse'. The nature of idioms is that they appear with too much irregularity to be classified as lexical bundles, and only manifest in natural spoken and written discourse with rarity, notwithstanding their more frequent employment in the register of fiction

(Biber 2006), while Simpson and Mendis (2003, cited in Biber 2006: 134) recorded the significant pragmatic functions in respect to the use of idioms in the classroom, finding that such expressions were uncommon, and more often than not manifested as short nouns or prepositional phrases.

The second noteworthy feature of lexical bundles is the fact that they are not typically complete from a grammatical perspective, as mentioned above, with Biber et al. (1999) finding that less than one fifth of lexical bundles featuring in conversation can be deemed 'as complete phrases or clauses', and that 'less than 5% of the lexical bundles in academic prose represent complete structural units' (Biber et al. 1999, cited in Biber 2006: 135); but rather the majority of lexical bundles span two distinct structural units, commencing at the boundary of a clause or phrase, and then the concluding words of the lexical bundle serve as the commencing components of the subsequent structural unit. 'Most of the bundles in conversation bridge two clauses (e.g. *I want to know, well that's what I*), while bundles in academic prose usually bridge two phrases (e.g., *in the case of, the base of the*)' (Biber 2006: 135).

Oakley (2002) refers to the lexical phrase category of 'sentence builder' as being non-continuous and with significant variations, enabling the framing of lengthy and challenging sentences, and therefore any non-native author in English would certainly require a clear understanding of those **versions** which are (or are not) acceptable, to engage with lexical phrases productively. The *Longman Dictionary of Language Teaching and Applied Linguistics* (2010) defines lexical phrases as those phrases and patterns that have recurred to the point where they have become assimilated into a language through regular use.

One particular approach to exploring classroom pedagogy has been through the consideration of the function of lexical phrases, chunks and

idioms that are somewhat longer (Biber 2006), with such studies representing an increasingly popular research theme that centres on how those multi-word expressions that have been prefabricated are typically employed. The study of multi-word sequences has been conducted under numerous 'rubrics, including "lexical phrases", "formulas", "routines", "fixed expressions", and "prefabricated patterns" (or "prefabs")' (Biber 2006: 133), with the approaches each specifying the focus of the research through differing nomenclature, to thus offer a range of viewpoints regarding multi-word sequences in language use. Some empirical investigations have focused on idiomatic multi-word sequences, while others have positioned their lenses on those multi-word sequences that are not idiomatic, as asserted by Biber (2006), who considers those multi-word sequences, or lexical bundles as he refers to them, that most frequently manifest in a particular register as focusing specifically on the pedagogical registers of classroom tuition and coursebooks, while creating an operational framework with which to analyse the bundles. He then carries out a comparison of those lexical bundles found across the entire spectrum of registers employed in university contexts, namely 'classroom management, office hours, study groups, service encounters, course management writing, and institutional writing', before conducting a final comparison of lexical bundles in use between academic fields in respect to their inclusion in coursebooks (Biber 2006: 133).

The issue of nomenclature indeed becomes clear in the case of considering alternative approaches to addressing the relationship that exists between lexis and grammar, as even prior to the advent of digitised corpora it had become apparent that employing syntactic rules to describe sequences of words would not be as comprehensive as when those same rules were applied to individual words, and therefore a specific terminology has been employed to define phenomena that are unable to be clearly described

through using grammatical analyses that tend to consider words as being individual in terms of meaning, for example, 'lexical phrases, idioms, fixed phrases, prefabricated phrases, formulae, institutionalised collocations, and compounds' (Mahlberg 2005: 20).

Such terminology, as Mahlberg (2005) observes, highlights more dovetailed relationships between words in some sequences than in others, together with the inability to always depend on a straightforward syntactic connection between individual words and meanings. The range of nomenclature describes the varying strategies for filling the explanatory void resulting from adopting the syntactic perspective, that is, the grammatical analysis for collocation that treats words as independent units of meaning. Mahlberg (2005) asserts that an alternative approach to considering co-occurring words that may not be analysed through the typical syntactic methods would be via the discourse function, as conventional commonly utilised phrases can be connected with certain contexts within the social environment, such as through greetings or formal introductions. The potential of corpus linguistics to introduce value and insight to the junction between lexis and grammar can manifest variously, such as awareness of the atypical nature of intense idiomatic expressions, where data resulting from corpora can offer additional insight into how such phrases can be found in texts (Mahlberg 2005). Moreover, Biber et al. (1999, cited in Mahlberg 2005: 20–1) reveal that certain idiomatic expressions may be found on occasion in fiction, but then may be difficult to locate in other registers.

Corpora offer users the potential to query and receive data that can be employed as evidence of the weaknesses present in the existing understanding of the variance and flexibility of idiomatic expressions, although while the awareness raising has value, offering greater clarity to the linguistic landscape will not result in significant shifts in the theory of

linguistics if the nomenclature remains unchallenged, as to discuss ‘idioms’, ‘fixed phrases’ and ‘formulae’ is to grasp at finite categories while suggesting an inflexibility to a system that in fact features rather adaptable permutations (Mahlberg 2005: 21).

5.5.2 Formulaic sequences

One contrast that can be found between the spoken and written forms of English can be said to be the positioning of what Mauranen (2004) refers to as ‘prefabricated elements (or gambits, formulae, lexical phrases – these semi-fixed, semi-formulaic expressions go under many names and definitions but are roughly comparable)’ (Mauranen 2004: 95–6), that have a firm presence in the communicative language teaching classroom, and more generally in the tuition of speaking skills; **moreover, formulaic sequences are of interest to learners of English since they offer invaluable chunks of language that once acquired, can serve to promote fluency and accuracy.** However, the teaching of such prefabricated elements has been said to be informed by the intuition of materials’ writers and established tradition as opposed to compelling data, and that consequently it has been revealed that a proportion of the well-established patterns that have traditionally formed the foundation of the development of speaking skills do not actually feature in natural speech (Mauranen 2004). Mahlberg (2005) states that the term ‘prefabricated phrases’ suggests the production of language from a psycholinguistic perspective that is pertinent to linguistic pedagogy.

Formulaic sequences can be understood and stored as a unit, without the initial need to comprehend the nature of the internal structure (although later analysis can enable the productive use of internal elements), while formulaic sequences can also be formed from smaller units and then archived as whole for later use (Longman Dictionary of Language Teaching and

Applied Linguistics 2010). The reference term of formulaic sequences contains a broad phraseological scope that is employed to indicate a wide range of connected concepts inclusive of 'chunks, collocations, fixed expressions, formulas, idioms, lexical phrases, prefabricated patterns, ready-made utterances, and routines', and although there has been considerable acknowledgement of the use of formulaic sequences in language use, its inherent diversity has created challenges in arriving at a final definition (Yoshitomi 2006: 203). Formulaic sequences can feature a variety of lengths, as well as a range of functions of a linguistic and social interaction nature, they can be rigidly fixed or allow other suitable words or phrases to be inserted within them at particular points, and while they may be distinct in respect to their level of denotational certainty; they have been described as a pattern of words or other components that are apparently prefabricated, and that are archived and recalled intact from the memory when required, as opposed to being subjected to analysis through the grammar of language (Wray 2000, cited in Yoshitomi 2006: 203). Flowerdew (2008) reports that learners were not aware of part-formulaic phrases that could be used as an alternative by dropping in a word or phrase to complete their intended meaning, and were seen to be trialling bespoke lexical phrases which could not withstand the scrutiny of realising native-like conformity, regardless of them being grammatically sound. Regarding the ability to insert words or phrases into formulaic sequences at particular points, the *Longman Dictionary of Language Teaching and Applied Linguistics* (2010) refers to these as lexicalised sentence stems.

Schmitt and Carter (2004, cited in Yoshitomi 2006: 203) highlight that despite the challenges of establishing concrete criteria through which formulaic sequences can be defined, there is the potential to describe some of their typical features, namely that they are (i) archived in the memory and

processed as a whole; (ii) produced through speech with greater fluency and clear intonational flows (Peters 1983, cited in Yoshitomi 2006: 203); and (iii) processed with increased efficiency due to their collective form, in contrast to alternative attempts to create the identical sequence of words in natural language use (Pawley & Syder 1983; Kuiper 2004, cited in Yoshitomi 2006: 203). Formulaic language is said to feature a number of functions, inclusive of preserving valuable resources required for processing and improving fluency, and enabling certain functions for interactional purposes (Richards & Schmidt 2010).

Of close related interest, but of potentially elevated importance is the evidence arising from the analysis of corpus whereby the overall meaning is not found in singular words but through phrases or collocations (Stubbs 2001, cited in Sealey 2009: 47), while a less conservative stance on this aspect has been embraced by a number of corpus linguistics, and therefore resulting from contrasts between the viewpoints of theory and analysis, one individual term to refer to the variety of possible longer phrases is lacking, with authors thus finding themselves with their attention focused on such terms as “multi-word items”, “formulaic sequences”, “formulaic language”, “lexical bundles”, “lexical phrases”, “prefabs”, and “chunks” (Sealey 2009: 47).

5.6 Key clusters

Baker, Hardie and McEnery (2006) refer to the term clusters as describing any such groupings of words that are in a sequence, such as the terminology employed in WordSmith Tools, as well as being a group of texts with linguistic characteristics that can be said to share statistical similarity (see Chapter 6, Section 6.6). Moreover, they define cluster analysis as a ‘multivariate’ method that enables the automated statistical querying and return of categories, and therefore the technique can be employed to measure

the extent that texts do or do not share similarity within the confines of the parameters determined by the researcher (Baker, Hardie & McEnery 2006: 34). With such methods offering valuable capability, Oakes (1998) adds that within the field of corpus linguistics a range of unique characteristics inclusive of 'case, voice or choice of preposition within a text' (Oakes 1998, cited in Baker, Hardie & McEnery 2006: 34) can be grouped or clustered together to evaluate and understand the use of such elements between differing registers, genres or authors.

5.7 Semantic prosody

With the initial investigations into semantic prosody beginning in the latter stages of the 20th century, exploring the sense of meaning generated via the characteristics of collocates that are connected in some way to the root node, a variety of titles have been coined to define this linguistic feature, and which have included the terms 'semantic preference and evaluative prosody' (Louw & Chateau 2010: 756).

Louw and Chateau (2010) conducted a study of semantic prosody-associated work, inclusive of investigating contemporary findings that cite the presence of semantic prosody in corpora that have been developed for a specialist purpose. They assert that semantic prosody is not exclusively characterised by being positive or negative, but rather that it can be moderated through particular contexts of a scientific or non-personal nature, although they also find that the core polarity can rapidly dominate in texts where strong debate features.

Louw and Chateau (2010) provide some historical context by reporting that the linguistic feature that is now referred to as semantic prosody was originally highlighted by John McHardy Sinclair in 1987, with Louw (1993) being credited for initially employing the term in a published article, although

relaying this merit back to Sinclair as the scholar who had originally brought the phenomenon to light. In Sinclair's work, the negative prosody of the phrasal verb *set in* is considered, and in particular the characteristics of the subjects, and the fact that it typically signposts disagreeable conditions or states (Louw & Chateau 2010: 756). Sinclair continued his investigation into the phenomenon in 1991, in part exploring semantic prosody in the context of 'Corpus, Concordance, Collocation', and exploring the negative prosody of an additional verb, *happen*. An in-depth discussion on the concept of semantic prosody, as presented by Louw, is also offered by Milojkovic (2013).

A further study of semantic prosody was carried out by Xiao and McEnery (2006), where they define the term as having 'collocational meaning' that highlights the recurring relationships that exist and connect the node word to other words with which it features with atypical frequency, and defining the resulting meaning from the collocation and collaboration between the node and associated words as semantic prosody, or 'a form of meaning which is established through the proximity of a consistent series of collocates' (Xiao & McEnery 2006, cited in Louw & Chateau 2010: 756).

Louw (2000) advances this description, stating that the semantic prosody connected to a word results from the semantically considered coherence the collocates create when viewed collectively, and that this perspective underscores the challenges of intuitively identifying the phenomenon, while additionally contributing to the 'misunderstandings and misinterpretations that have cluttered the debate since the phenomenon was first brought to light' (Louw 2000, cited in Louw & Chateau 2010: 756). The primary function of semantic prosody is thus defined by Louw (2000, cited in Louw and Chateau 2010: 756) as a vehicle that enables a speaker or author to express their position towards particular practical scenarios.

In contributing to the discussion on defining the nature of semantic prosody, Zhang (2010) underscores three identifiable and specific characteristics: (i) 'functionality', which relates to the speaker or author selecting those lexical items that will lead to the generation of intelligible sentences; (ii) 'linguistic choice', where the selection and association of collocates is actively and carefully considered; and (iii) 'communicative purpose', in order to express 'attitudinal meaning' (Zhang 2010: 190).

Baker, Hardie and McEnery (2006) refer to the term semantic preference, which while sharing similarities with the notion of semantic prosody, defines 'the relation, not between individual words, but between a lemma or word-form and a set of semantically related words' (Stubbs 2001, cited in Baker, Hardie & McEnery 2006: 144). An example of semantic preference in relation to words is presented, where through querying the British National Corpus the word *rising* has a tendency to collocate with words associated with employment and income, while in the case of phrases, for example, the word *glass* has a tendency to co-occur with words associated with drinks (Baker, Hardie & McEnery 2006). **Bednarek (2008) re-examines the concept of *semantic preference* and the notion of semantic prosody, suggesting that they be treated 'as two types of the same *collocational phenomenon*', since the 'two types of collocation are ... very similar, differing only in degrees of "generality", and frequently occur together' (Bednarek 2008: 121).** It can thus be said that the nature of semantic preference is associated with notions of 'collocation and colligation', while placing its lens on 'a lexical set of semantic categories' as opposed to individual words, or rather a group of grammatical words that have some relation (Baker, Hardie & McEnery 2006: 144). Although being related to discourse prosody, the distinction between the two phenomena can be challenging to describe (Baker, Hardie & McEnery 2006), and while Stubbs (2001, cited in Baker, Hardie & McEnery

2006: 144) points to the impact of the open-endedness of the collocate list, whereby it might be feasible to produce a list of all those words that can collocate with *drink*, thus suggesting semantic prosody, a category that contains more disagreeable conditions or states might more accurately imply discourse prosody.

In their discussion on semantic prosody, Baker, Hardie and McEnery (2006) consider the notion that in language use words can collocate with particular semantic sets in addition to specific words, using the example of 'the word *hair* [that] may collocate with semantic groups such as length (*long, short*) and colour (*red, blonde, black*)' (Baker, Hardie & McEnery 2006: 145), while they assert that investigation of concordances can facilitate the identification of semantic prosodies.

Where semantic preference has the ability to describe the particular semantics of a word or indeed a phrase, semantic prosody reveals the essential values that the speaker or author has embedded in their discourse, and while sharing similarities with connotation, the application of semantic prosody is to more than the individual word, extending to include both the node and its related collocates, to either assume negative or positive semantic prosody depending on the nature of the collocates (Flowerdew 2009).

The objective of analysing those semantic prosodies that share a relationship with the lexical items found within a corpus offers the researcher the opportunity to gather contextual information that can prove valuable for those authors who are attempting to acquire and develop their skills in a particular field of writing (Flowerdew 2009), with Tribble (2000, cited in Flowerdew 2009: 333) pointing out that while such information is beginning to feature within dictionaries, the educational and enlightening process of the contextual analysis of words and their semantic prosodies could prove to be a

more profound learning experience, where the storage and recall of such information occurs with greater depth and clarity.

In consideration of genetic criticism, Wolf (2014) claims that the stance is critical in emphasising the perspective of literary work 'as primarily a reflection of the author's life and times', with the term arousing 'images of heredity and lineage from the author to his or her work' (Wolf 2014: 23). The investigation of an author's works extends far beyond the limits of the biographical confines, as such critics are concerned by their very nature with causality, and so where an author creates a poem, and that author's work is symbolic of the era in which he/she was placed, then in order to fully understand the author's craft we must become intimately familiar with that period, inclusive of 'the author's entire political, social, and intellectual milieu' (Keesey 1987: 11, cited in Wolf 2014: 23).

Flowerdew (2013) also points towards the meaning connections which words inherently carry due to their common collocations with groups of other words that have a semantic relationship, with semantic prosody also conveying a more practical meaning in describing or alluding to the speaker or author's attitudinal feelings towards the subject of the moment, be they positive or negative; however, it is pertinent to remain mindful that semantic prosodies cannot be intuitively accessed, and must be established through careful and targeted engagement with a corpus (McEnery & Hardie 2012, cited in Flowerdew 2013: 164).

5.8 Conclusion

In this chapter I have explained and discussed the primary analytical techniques of corpus linguistics. These are the techniques that I apply in Chapter 6 with the aim of revealing more about the works of Dickens by means of corpus stylistics. In so doing, I will discuss how such findings can be

utilised to facilitate the comprehension of the text besides other possible pedagogical applications when teaching Dickens's works for non-native readers.

Chapter 6 Analysing the Charles Dickens Complete Corpus

6.1 Introduction

Following the discussion regarding the construction and description of Charles Dickens's Complete Corpus (DCC) in Chapter 4, this chapter shifts its attention to the potential pedagogical applications regarding the analysis of its data. The lens of the analysis contained within this chapter will focus in particular on the quantitative exploration of Dickens's works by considering the *wordlists*, *keywords*, *text collocations*, *lexical bundles/clusters*, and finally, *semantic prosody*. Besides identifying the frequencies of occurrence in the DCC, the role that such findings can play in studying and learning to appreciate Dickens's works will also be explored, in addition to examining the manner in which these findings (relating to the frequencies in the DCC) can be utilised to facilitate the process of reading and then comprehending the texts, based on that evidence emerging from the frequency analysis. Analysis of the individual lexical items, phrases, collocations, lexical bundles or semantic prosody will be performed so as to demonstrate the pedagogical potentiality of such findings, and particularly in the context of non-native readers. As Bateman (2008) asserts, the value of employing corpus linguistics' techniques is in 'finding the patterns' in a corpus that can contain millions of words (Bateman 2008: 250), as is the case with the DCC. Thus, it is essential to harness tools that can automate such lengthy processes. Baker (2006a) indicates towards frequency being regarded as a fundamental concept that supports the process of corpus analysis, with frequency lists offering valuable potential as an initial area of focus when analysing all types of corpora due to (i) their inherent nature as one of those elementary tools employed by the corpus linguist, and (ii) their potential to reveal a range of insightful

characteristics. He additionally explores the manner in which researchers can exploit frequency lists in order to explore certain regions of a corpus, asserting that when the respective corpus is engaged with, such frequency data can map the usage of a word or phrase from a sociological perspective in order to shine light on how it is utilised in particular contexts (Baker 2006a).

In addressing the question of why discourse analysis might benefit from measuring frequency, Baker (2006a) claims that language does not merely manifest in some haphazard manner, but rather that certain words have a tendency to co-occur in association with others, and that this phenomenon manifests with a regularity that can be predicted with surprising accuracy. He **mentions** that languages are governed by a plethora of rules which dictate what may or may not be written or spoken in a particular context or point in discourse, and that with the use of language being a compromise between free choice and defined patterns, the nature of frequency has major relevance as an author or speaker's record of language use in a unique or atypical fashion offers the potential to garner insights into their purpose, regardless of whether their linguistic choices have been made either consciously or unconsciously. Since there are patterns in language use that can be illuminated through exploiting language corpora, assisting learners in familiarising themselves with such usages can be of interest as they are typically advised to acquire the most frequent words related to a specific genre or the language in general. In other words, the learners' focus should be equally directed towards both the acquisition of the most frequently occurring words and their usages, which can be reasonably identified through engagement with corpus techniques. Non-native speakers at advanced levels should thus focus their attention on acquiring lexical items. Richards and Schmidt (2010) **point out** that such a *lexical approach* to language acquisition is grounded in the belief that the fundamental components of

teaching and learning comprise of ‘words and lexical phrases, rather than grammar, functions or other units of organization’, and that with lexicon being considered to have a central role in the organisation, learning and teaching of language (as opposed to grammar), its consideration is fundamental to ‘syllabus design, course content, and teaching activities’ (Richards & Schmidt 2010: 355).

6.2 *The wordlist frequency of the DCC*

In the analysis, two software tools will be engaged with in order to extract the required information from the DCC: AntConc 3.4.4w by Anthony (2005), and the WordSmith Tools 6.0 suite by Scott (2011). The **DCC** consists of the sub-corpora presented in Table 6.1.

Table 6.1 The DCC sub-corpora and word counts

File no.	Work	Word count
01	Sketches By Boz, Illustrative of Every-Day People	261,445
02	The Pickwick Papers	313,634
03	Oliver Twist	165,925
04	Nickolas Nickleby	332,526
05	The Old Curiosity Shop	222,511
06	Barnaby Rudge	260,428
07	American Notes For General Circulation and Pictures From Italy	178,170
08	A Child History Of England	164,839
09	Martin Chuzzlewit	346,741
10	Christmas Books	158,948
11	Hard Times and Other Stories	139,151
12	Dombey and Son	363,605
13	David Copperfield	363,480
14	Bleak House	363,839
15	Little Dorrit	345,606
16	Christmas Stories From ‘Household Words’ And ‘All Year Around’	358,490
17	A Tale Of Two Cities	138,366
18	Great Expectations	188,900
19	The Uncommercial Traveller	145,938
20	Our Mutual Friend	333,800

21	The Mystery of Edwin Drood – And Master Humphrey’s Clock	146,906
22	Reprinted Pieces – Sunday Under Three Heads	150,387
23	Miscellaneous Papers – Plays and Poems	353,731
24	Letters and Speeches	406,520
Total tokens of the DCC		6,202,886

Through employing the AntConc 3.4.4w corpus analysis software, a wordlist was generated from the DCC text files. As discussed in Chapter 5 (see Section 5.2), the wordlist merely represents a list of all those word types that feature in a corpus, together with data describing their frequencies, which has the potential through AntConc 3.4.4w to be ordered in respect to (i) frequency, (ii) alphabetical order, or (iii) by the end of the word. While the complete token count of the DCC is 6,202,886 tokens, the obtained raw wordlist contains merely 47,400 word types, that is, less than 1% (0.764%) of the DCC’s entire tokens. Reflecting a low type-to-token ratio, this finding clearly underscores the volume of repetition that features in the DCC. As pointed out by Baker (2006a), any corpus that is found to have a low proportion of word types to tokens will by definition feature significant repetition (whereby individual words are repeated with great frequency), while any corpus that returns a high proportion of word types to tokens indicates towards a broader language usage. Despite the fact that these ratios of word type to token can be supportive when considering comparatively smaller text files, as the size of the corpus increases the ratio of word type to token will invariably decrease, as those grammatical words that feature with high frequency will repeat regardless of the corpus’s actual size. Therefore, large corpora invariably feature low ratios of word type to token and intra-comparisons can become problematic, although the ratio of word type to token only offers a glimpse into the complexity of lexis, and hence additional study is required (Baker 2006a). Notwithstanding these challenges, insight can be gained through comprehending the manner in which elementary

frequency-based wordlists can facilitate our understanding of the harmony and discord between two textual sets.

It is worth underscoring at this point that the wordlist of the DCC is a raw one, with the definition of a *raw wordlist* in this context being that the DCC contains, for instance, a number of function words such as *the, and, of, to*, etc. Furthermore, it features auxiliary verbs along with their contractions (e.g. *are/aren't* and *can/can't*), while additionally comprising of prepositions, conjunctions, determiners and pronouns (some of which are archaic, such as *thou, thee*, etc.), and numbers primarily written out in full, as opposed to their numerical counterparts (e.g. *seven* vs 7). The wordlist also contains a number of non-conventional/non-standard spellings of some words that were intentionally spelt as such; for example, to represent a dialectal pronunciation or allude to the level of education, and to a hint towards a social group, besides other functions. The DCC wordlist can be refined by excluding function words and the characters' proper names. Nation (2001), for instance, has prepared a list comprising of a total of 320 word types that he considers to be function words (see Appendix 1.1), and which will be harnessed in order to refine the DCC wordlist. Moreover, Hawes (2002) provides for a more detailed inclusion of Dickens's characters, although this does not include what he characterises as the non-fictional works of Dickens, that is *American Notes, Pictures from Italy, The Life of our Lord, A Child's History of England, The Lazy Tour of Two Idle Apprentices* 'and later collections of Dickens's uncollected writings' (Hawes 2002: xxv). In the case of the DCC, the corpus features the proper names of characters that appear in all Dickens's works (i.e. both his fiction and non-fiction), besides other real names that are found in his personal letters. In *The Oxford Companion to Charles Dickens*, edited by Schlicke (2011), an 'Alphabetical List of Characters' is provided that

includes most named characters who appear in the novels, Christmas books, stories, sketches, and plays. It includes minor figures as well as main characters, and major generic figures such as Bishop, Bar, and Physician. It does NOT include characters from Dickens's journalism ... characters from *The Mudfog Papers*, *The Uncommercial Traveller*, and *Reprinted Pieces* are likewise omitted.

(Schlicke 2011: 622; capitals in original)

I created a special list of Dicken's characters that merge the lists extracted from Hawes's (2002) *Who's who in Dickens* and Schlicke's (2011) *The Oxford Companion to Charles Dickens*, which produces 1,945 proper names with no **repetition**. Through considering the above-mentioned list, the DCC contains 1,879 character names, which through excluding them from the wordlist can further reduce those word types that may offer reduced significance in respect to stylistically interpreting the text. As Baker (2006a) suggests, function words (also referred to as grammatical words) have a tendency to remain unaltered by linguistic invention as it is not common practice to create innovative new pronouns or conjunctions.

An additional rationale for excluding function words in order to direct the focus towards the content words is that the function words exclusively feature with high frequency. This can be demonstrated in Table 6.2 below, where the twenty most frequently occurring words are presented. For the analysis I used WordSmith Tools 6.0 (Scott 2011) here, as besides the frequency numerical value it provides the frequency percentage of each word and its range in the 24 text files of the DCC.

Table 6.2 The twenty most frequent words in the DCC

No.	Word	Frequency	%	Text files
1	the	317,343	5.16	24
2	and	221,710	3.61	24
3	of	171,657	2.79	24
4	to	164,867	2.68	24

5	a	138,199	2.25	24
6	in	110,151	1.79	24
7	I	109,514	1.78	24
8	that	81,015	1.32	24
9	it	76,237	1.24	24
10	he	71,449	1.16	24
11	his	70,754	1.15	24
12	was	66,621	1.08	24
13	you	61,769	1.00	24
14	with	57,759	0.94	24
15	as	51,225	0.83	24
16	for	44,380	0.72	24
17	at	42,640	0.69	24
18	had	42,509	0.69	24
19	is	38,944	0.63	24
20	her	38,878	0.63	24
Total		1,977,621	32.14%	

It can be seen that all the twenty most frequently occurring words in Table 6.2 are function words that can be found in Nation's (2001) list of function words. These function words represent 1,977,621 of the total tokens of the DCC (6,202,886), representing 32.14% of the entire corpus of Dickens. In order to produce a list of Dicken's headwords focusing solely on the lexical words (content words), I will thus exclude both the function words and the proper names. AntConc 3.4.4w features a number of options that can be harnessed when generating wordlists, one of which is to ignore a certain set of words by exploiting the '*Use a stoplist below*' feature in the Tool Preferences – word list. Through employing Nation's (2001) list of function words and the special list of Dicken's characters that I generated by merging the lists of Hawes (2002) and Schlicke (2011), I created a *stoplist* that features the function words and the characters' proper names. With the intention of refining the Dickens's Word List (which contains the headwords, their family members and proper names), I thus created a list that can be referred to as 'The Dickens's Lexical Words List'. Lexical words (also known as content words) convey greater information and meanings than their functional (grammatical)

counterparts (Gee 1999). This list additionally features the frequency of each lexical item in the entire corpus, to be discussed later when I address the stylistic choices made by Dickens. The list consists of a total of 25,445 words and can be accessed in full by visiting <http://archive.org/details/DCCLexicalWordsList>, while Table 6.3 presents a sample of the twenty most frequent lexical words occurring in the DCC.

Table 6.3 The twenty most frequent lexical words in the DCC

Lexical word	Frequency	Range
say	47,048	24
mister	35,292	24
go	19,594	24
know	18,054	24
look	16,550	24
see	15,915	24
make	15,783	24
take	14,522	24
think	14,335	24
time	13,818	24
hand	11,510	24
great	11,117	24
like	10,082	24
day	9,809	24
dear	8,519	24
well	8,455	24
head	7,983	24
way	7,696	24
eye	7,693	24

It is interesting to notice that less than **0.01%** of the total number of the DCC tokens (i.e. the raw wordlist, which is 47,399 in total) were employed by Dickens to produce such a significant number of works. An attempt to explore this phenomenon can be achieved by responding to the question of how Dickens constructs the meanings in his work. The use of the term *constructing the meaning* here **implies** that Dickens, as per other writers, has a pool of materials (i.e. lexicons at the semantic level, grammar at the syntactic

level, sounds at the phonological level, etc.) from which he can select and then construct meanings or create concepts. The investigation can also be applied per se at each of the linguistic levels. For instance, at the semantic level the investigation can extend to questioning the relation between the semantic senses that present themselves across words (i.e. the relationship between lexical items in respect to synonymy, antonymy, hyponymy, and polysemy). Then, this can be further explored in terms of the pragmatic employment of the lexical items. This manner of investigation has its focus centred at the intertextual level, where the *lexical semantics* and *grammatical semantics* are explored. Cruse (2006) defines lexical semantics as being

the systematic study of meaning-related properties of words ... how best to specify the meaning of a word; **paradigmatic** relations of meaning such as **synonymy**, **antonymy**, and **hyponymy**; **syntagmatic** relations of meaning, including **selectional restrictions**; structures in the lexicon such as **taxonomic hierarchies**; change of word meaning over time; and processes of meaning extension, such as **metaphor** and **metonymy**.

(Cruse 2006: 95; bold in original)

It can be noted here that both lexical semantics and grammatical semantics 'may exclude aspects of meaning treated under **pragmatics**' (Cruse 2006: 95; bold in original), which can be considered as being the intratextual level of investigating the construction of meanings in texts. **The intertextuality, as Fischer-Starcke (2010) suggests, is concerned with the 'linguistic patterns that occur only within one text and which contribute to its meaning. These patterns are called intratextual references and form an intertext within one text' (Fischer-Starcke 2010: 35). On the contrary, 'intratextual references are not recognized and decoded by all recipients of a text' due to the fact that these 'Factors, such as a receiver's individual textual competence, the situational context of reception and the receiver's background knowledge**

[vary from one recipient to another, while they] influence the interpretation of a text' (Fischer-Starcke 2010: 35).

With the intention of retrieving the lexical lemmas from the raw DCC wordlist, two measures were applied: (i) exclusion of the function words using Nation's (2001) list, and then (ii) lemmatising the entire list using the English Lemma List v.1 compiled by Someya (1998), which 'currently contains 40,569 words (tokens) in 14,762 lemma groups' (Someya 1998: para. 1). These steps resulted in 2,660,200 word tokens due to the removal of the function words, which were then lemmatised to produce 34,307 word types. The top twenty most frequently occurring lexical lemmas are presented in Table 6.4 below.

Table 6.4 The twenty most frequent lexical lemmas in the DCC

No.	Lemma	Frequency	Lemma word forms
1	say	45,900	said (33,216), say (9,115), says (3,569)
2	go	19,487	go (6,041), goes (925), going (4,601), gone (2,615), went (5,305)
3	know	17,624	knew (2,772), know (11,031), knowing (807), known (1,844), knows (1,170)
4	look	16,495	look (5,178), looked (5,207), looking (5,038), looks (1,072)
5	see	15,804	saw (3,517), see (7,983), seeing (930), seen (3,156), sees (218)
6	take	14,455	take (5,275), taken (2,358), takes (600), taking (1,870), took (4,352)
7	time	13,727	time (11,407), timed (33), times (2,287)
8	make	13,625	made (7,930), make (5,045), makes (650)
9	hand	11,451	hand (7,247), handed (359), handing (91), hands (3,754)
10	great	10,478	great (9,127), greater (827), greatest (524)
11	like	9,929	like (9,346), liked (358), likes (139), liking (86)
12	day	9,332	day (7,509), days (1,823)
13	dear	8,377	dear (7,731), dearer (43), dearest (521), dears (82)
14	think	8,256	think (6,953), thinking (1,010), thinks (293), thought (4,544), thoughts (966)

15	head	7,956	head (6,909), headed (348), heads (699)
16	way	7,666	way (7,146), ways (520)
17	eye	7,328	eye (1,876), eyeing (118), eyes (5,334)
18	night	7,316	night (6,988), nights (328)
19	give	7,250	gave (1,961), give (3,011), given (1,212), gives (321), giving (745)
20	leave	7,005	leave (2,073), leaves (551), leaving (623), left (3,758)

Adolphs (2006) suggests that '[w]ordlists of individual texts can highlight items that are characteristic for a particular domain, author or text-type' (Adolphs 2006: 27). Through considering the most frequent lexical lemma *say* with a frequency of 45,900 (33,216 for *said*, 9,115 for *say* and a frequency of 3,569 for *says*), the following point can be noted: The works of Dickens include a fair amount of reporting. Dickens did not narrate the majority of his works in the *first person narrator* style, **despite** it being more typical for the first person narrator in novels to repeatedly employ the first person pronoun *I*, for instance. While discussing the mode of narration Davis (1999), for example, confirms that Dickens 'used first-person narration in only three of the novels' (Davis 1999: 136): *David Copperfield*, *Great Expectations* and partially in *Bleak House*. Despite the ability to easily recognise intuitively whether the novel has been written in the first person narrative, I examined the fifteen novels of Dickens for the first person pronoun *I* usage in order to establish the concordance plot for the novels, as described in Table 6.5.

Table 6.5 The concordance plot for 'I' in the DCC

No.	Novel	Tokens	'I' Frequency
1	David Copperfield	363,480	13,467
2	Bleak House	362,839	9,506
3	Our Mutual Friend	333,800	6,786
4	Great Expectations	188,900	6,673
5	Little Dorrit	345,606	6,012
6	Martin Chuzzlewit	346,741	5,591
7	Nickolas Nickleby	332,526	5,046

8	Dombey and Son	363,605	4,861
9	The Pickwick Papers	313,634	3,926
10	Barnaby Rudge	260,428	3,320
11	The Old Curiosity Shop	222,511	2,591
12	Hard Times	105,663	2,226
13	A Tale of Two Cities	138,366	1,991
14	Oliver Twist	165,925	1,958
15	The Mystery of Edwin Drood	97,850	1,894

Through Table 6.5 it can be clearly noted that the first person pronoun 'I' appears with great significance in *David Copperfield* (13,467 instances), *Bleak House* (9,506) and *Great Expectations* (6,673), which are all narrated in the first person narrative mode. What is also interesting is to understand why the pronoun *I* appears 6,786 times in *Our Mutual Friend*, although it was not narrated in the first-person perspective. This analysis offers a clear indication that the pronoun *I* has been frequently employed in these three of Dickens's novels, which can be understood when we refer forward and discover that these three novels are written mainly or partially in the **homodiegetic** narrative style.

As for the lexical usage of *say*, it can be underscored that this lexical item is employed for different purposes inclusive of expressing views and opinions, providing information or examples, revealing thoughts and feelings, and in speech and repeat words, in addition to demonstrating that someone has the authority to decide on a particular issue. From continued investigation of the word *say*, an attempt can also be made to discover what can be classified as spoken phrases. Moreover, searching the DCC for *say* + adverb (*say *ly*) resulted in 116 instances of adverbs that could be considered and tested for teaching purposes. These examples offer some indication of the extent of those applications that can be studied through the wordlist of a given corpus. The next section will proceed to address the keywords in the DCC.

6.3 Keywords

McEnery and Hardie (2012) indicate that through comprehending and describing the frequency data that emerge from a corpus in greater depth, the corpus linguist will refer to statistical measurements that enable the potential to reposition the direction of analysis from being merely descriptive towards being able to actually test the relevance and implications of the data that emerge. Furthermore, they assert that with the majority of textual features that we may wish to test being 'subject to a certain amount of "random" fluctuation', then 'significance tests' can be employed to evaluate the potential for a corpus result to have emerged coincidentally, and despite the fact that these results can be utilised to signpost where further study may be required, 'if there is a 95 per cent chance that our result is *not* a coincidence', then it can typically be deemed to be of significance (McEnery & Hardie 2012: 51; italics in original).

Scott (1997) employs the nomenclature of keywords to indicate lexical items that co-occur either with considerably elevated or decreased frequency in a text (these being referred to as positive or negative keywords, respectively), with the comparison being drawn between the target works and a reference corpus (cited in Adolphs 2006: 44). Keywords can be defined within the spectrum of word frequency lists that have been sourced from the corpus of focus and the reference corpus, and then a statistical comparison made, whereby the items identified in the target corpus are then measured against those in the reference corpus and the difference calculated to determine its statistical significance (Adolphs 2006). While these statistics both consider the distance between the resulting frequencies of two items and those which might be expected to arise, where the difference is substantial the potential then arises to conclude that the relationship between the two items is not one that has formed by chance, but rather that there will be additional

factors of influence; therefore, the process results in words that both reflect and fail to reflect the corpus of focus (Adolphs 2006). Keywords have additionally been described by Scott (1997) as those words that feature with a frequency that is considered to be unusual rather than high, when compared against a reference corpus. In Scott's (1997) study the following keyword identification process is followed: (i) a list of keywords is generated from the **comparison** reference corpus, inclusive of all the keywords and frequencies that feature; (ii) the same procedure is then applied to the target text; (iii) a comparison is drawn between the keyword list arising from the target text, and that from the reference corpus **Scott (1997)** highlights that while the 'keyness' of the keyword is acquired through chi-square statistical analysis, the primary characteristic to be comprehended is the notion of the keyword's relevance (**i.e.**, if the word appears with frequency then it will be deemed to be pertinent to the text); before finally, (iv) once all the keywords have been identified, they are ordered in respect to their importance.

In the light of the notion of keywords, the selection process followed in order to determine which reference corpus will be employed for comparative purposes is vital, and which Adolphs (2006) suggests can be seen in the case, for example, of opting to compare transcripts resulting from medical consultations with a reference corpus that contains merely written texts, where the contrasts between the two productive forms of language use are likely to negatively impact on the validity of the keywords that emerge from the comparison between two contrasting corpus. Adolphs (2006) observes that 'the analysis of keywords, i.e. those words that occur with a significantly higher or lower frequency in a text in relation to another body of text, can be useful to establish an initial characterization of a particular type of discourse' (Adolphs 2006: 27). In order to carry out this comparison, the concept of a reference corpus demands a brief discussion. Baker (2006a) refers to a

reference corpus as that which the majority of purists would term as merely a corpus: it containing a body of words numbering in their millions, sourced from a broad range of texts, and being reflective of a specific language variety or genre, such as the British National Corpus (BNC) that features more than one hundred million spoken and written words, and the British English 2006 (BE06) Corpus that contains more than one million words of generally employed British English.

In order to generate a keyword list for the DCC, the wordlist of the BNC was selected as a reference corpus, together with the BE06 Corpus created by Paul Baker; the rationale behind the selection of these two reference corpora being the assumption that learners of English in the modern era are learning a contemporary variety of modern English which can be claimed to a great extent to be represented in these two corpora (i.e. the BNC and BE06). Baker (2006b) describes the BE06 Corpus as a corpus containing one million words that originate from general British English in its written form, employing the sampling frame utilised by the Lancaster-Oslo/Bergen Corpus (LOB) and Freiburg-LOB Corpus of British English (F-LOB), and which comprises 'of 500 files of 2000 word samples taken from 15 genres of writing', with the vast majority (82%) of the texts being published in the 2005–2007 period, while the remainder were published in the 2003–2004 and early 2008 periods. With the medial point of sampling being 2006, this year is thus reflected in the title of the corpus. The rationale for employing these two temporary British English corpora is to extract a keyword list that has been compared with the modern English usage, with the assumption being that modern learners of the English language are engaging with the contemporary English form.

The Log-Likelihood method was selected as the statistical measure through which to determine the significant lexical items in the DCC. The

significance value applied was as follows: 95th percentile; 5% level; $p < 0.05$; critical value = 3.84. This resulted in a keyword list that consists of 6,616 keywords. This list was reduced still further by creating a stopword list from 'The Alphabetic List of Characters' provided in *The Oxford Companion to Charles Dickens*, edited by Schlicke (2011). The keyword list was again reduced to 5,728 following the exclusion of the character names from the keyword list, while the final measure applied to the DCC keyword list was to lemmatise the list by employing Someya's (1998) English Lemma List, which condensed the keyword list down to 5,478 words (see Appendix 6.1 for the DCC keywords placed in alphabetical order). The keyword list can now be manually refined as it still features several names such as *Abel, Adelph, Norah*; a number of place names such as *Ohio, Niagara*; and several unconventional spellings, for example, *goin* (going), *arter* (after), *arternoon* (afternoon), *arterwards* (afterwards). The analysis of the keyword list can also reveal certain characteristics of the DCC through utilising a comparable corpora to facilitate in establishing the 'aboutness' of the corpora, as it would then be feasible to construct a specialised keyword list for Dickens's works that functions as a basis for reading said works.

The AntWordProfiler 1.4.0w by Anthony (2013) was also employed to generate a DCC word-family list bearing similarity to the Academic Word List by Coxhead (2000). The DCC Headword List is available at <http://archive.org/details/Appendix6.2HeadwordsWithFamilyMembers> without the family members of the headword, resulting in the list featuring 2,486 headwords in total. By employing the notion of word family introduced by Paul Nation (e.g. see Hirsh & Nation 1992), a word-family list of the DCC was created using Nation's Range programme, in association with the family members of each headword. The DCC Word-Family List contains approximately 102,753 words: the headwords and their family members. The

list still includes a number of unconventional spellings of some words (e.g. *tricklin* (trickling), *tryin* (trying) and *turnin* (turning)) which can be considered to be family members only in Dickens's texts. An example from the DCC Word-Family List is provided in Table 6.6.

Table 6.6 Example from the DCC Word-Family List

A
an
aback
abaft
abandon
abandoned
abandoning
abandonment
abandons
abase
abased
abasement
abases

6.4 Collocations

In their discussions on the merits of studying collocation, Webb and Kagimoto (2011) point towards the general consensus in the contemporary era that acquiring such knowledge represents a valuable resource for language learners, as increased awareness and understanding of collocation both facilitates enhanced accuracy while simultaneously scaffolding fluency and skills' development at the pragmatic level; therefore, collocation is considered to be a key player in supporting learner competency between languages, and is thus finding itself under the lens of the academic spotlight of both theoretical and pragmatic studies. By considering the collocations from a statistical perspective, one of the keywords in the DCC (*eye*) will be investigated through utilising WordSmith Tools 6.0.

Table 6.7 Top twenty collocates for 'eye' to the left of the node collocate

No.	Word	Relation	Texts	Total Left
1	needle's	12.033	5	5
2	kindling	9.625	5	6
3	twinkling	9.495	10	12
4	watchful	8.132	7	7
5	glistening	7.952	6	6
6	practised	6.182	8	10
7	bright	5.44	21	38
8	flaming	5.357	1	6
9	eager	5.334	5	6
10	naked	5.101	6	9
11	evil	4.971	8	11
12	keen	4.877	10	8
13	quick	4.225	10	14
14	closed	3.382	9	13
15	blue	3.279	9	10
16	grey	2.585	3	12
17	public	1.768	12	15
18	human	1.499	5	8
19	left	1.118	9	14
20	own	0.507	9	9

Through excluding the function words in this short list of collocating words, the words listed in Table 6.7 accrue with the keyword *eye* (i.e. *bright eye*, *left eye*, *public eye* ... etc). These twenty items presented in the table occur to the left of the node collocate *eye*. The column *Relation* indicates to the strength of the relation, that is, the strength that each collocate relates to the association between the node (search word) *eye* and its collocate. The *Total Left column* indicates to the frequency of each item in relation to the node.

These twenty items presented in Table 6.7 occur to the left of the node collocate *eye*, as can be noted in the *Total Left* column, besides occurring in additional spans in the table. By excluding the function words in this short list

of the collocating words, the words listed above accrue with the keyword *eye* (i.e. *bright eye, left eye, public eye ...* etc).

6.5 Lexical bundles

Ndsi and Basturkmen (2009) report that one method of identifying lexical bundles is via the adoption of empirical measures, as opposed to those reliant on intuition, whereby through the former approach the patterns of words that co-occur with frequency can be highlighted in a target text or corpus without the influence of syntactic boundaries or the ability of the words to function beyond the confines of the context in which they are found. Patterns or strings of words of particular specified lengths can be located, thus allowing for their meaning to be explored, with such strings of words being referred to variously as clusters, recurring word combinations, statistical phrases, lexical bundles and n-grams (Ndsi & Basturkmen 2009) – the *Longman Grammar of Spoken and Written English* considers lexical bundles extensively.

In reference to the unit of analysis, Schmitt (2007) points out that while the clause prevails as the analytical unit in the majority of grammars, interest now extends within and beyond the clause, reaching towards an increasing awareness of the role of lexical items; while with reference to Universal Grammar, Cook (1994) ‘observes that the acquisition of syntax is minimized and the acquisition of vocabulary items with lexical entries is maximized’, and that ‘[t]here is also a recognition of the lexical phrase as a significant unit in language development’ (Cook 1994, cited in Schmitt 2007: 834), such as in the cases where learners often commence with a reliance on established chunks of language before engaging in greater scrutiny of their component parts.

Schmitt (2007) moves forward by making the prediction that corpus studies will continue highlighting unusual collocational strings of a lexical

and grammatical nature which frameworks of a conventional and more descriptive type are typically unable to reveal, while pointing towards the opposing end of the continuum where the unit of analysis shifts to discourse, with the emphasis being placed on the manner in which grammatical meaning is reliant on the narrative of the text.

Despite text being a semantic rather than grammatical unit, its meaning can be understood through the manner in which something is written or said, namely the wording, and thus any personal interpretation of the meaning embedded within a text is not feasible without first possessing a working theory of grammar; therefore, the comprehension and illumination of meaning requires 'a discourse grammar' that is 'functional and semantic in its orientation, with the grammatical categories explained as the realization of semantic patterns', or the resulting orientation will be merely internal, with the judgements and insights formed not being grounded in the texts' 'situational and cultural environment' (Schmitt 2007: 834).

A significant body of contemporary research has centred on establishing the identifying 'lexicogrammatical features of the discourses' emerging from a range of academic subjects and disciplines (Schmitt 2007: 835).

In considering the broader discussion on the potential of lexical phrases and formulaic language to facilitate the tuition of academic writing, Oakey (2002) asserts that there is no consensus, with Lewis (1993, cited in Oakey 2002: 127) positing that those lexical phrases that have been pre-determined as being of acquisitional value may be introduced to learners in accessible scenarios, to be acquired and understood as a unit, and thus offer scaffolding towards the achievement of syntactic mastery.

Through following similar criteria suggested by Biber et al. (1999), three-word and four-word bundles will be extracted from the DCC. As Biber et al. (1999) state

[t]o qualify as a lexical bundle, a word combination must frequently recur in a register ... lexical sequences are counted as “recurrent” lexical bundles only if they occur at least ten times per million words in a register. These occurrences must be spread across at least five different texts in the register (to exclude individual speaker/writer idiosyncrasies).

(Biber et al. 1999: 992–3)

The rationale for application here in the DCC is in the context of those lexical bundles that occur at least five times in five different texts (ranges as in AntConc 3.4.4w Tools) of Dickens’s works.

The lexical three-word bundles of the DCC were extracted via AntConc 3.4.4w, where the total number of n-gram types was 100,946 and the total number of n-gram tokens 1,735,319. In Appendix 6.2, only the three-word lexical bundles that occur in the 24 files of the DCC are presented, as the entire result of the three-word bundles identified in the DCC would be too substantial to be included as an appendix.¹

6.6 Key clusters

It has been asserted by Baker (2006a) that an alternative approach to identifying words that occur with frequency in several groups of texts that are comparable but may have distinct application would be to place the focus on the key clusters of words, rather than the keywords, as through harnessing WordSmith Tools there is the potential to query for lists of clusters, as opposed to individual words. The resonance between these two lists can then be considered, in order to establish those ‘combinations of words [that] occur

¹ The entire result of the three-word bundles of the DCC can be obtained upon request.

more frequently in one text or corpus when compared with another: a list of *key clusters*', with WordSmith Tools enabling parameters to be established concerning the size of the cluster, as lower returns of key clusters typically result when larger clusters are queried (Baker 2006a: 140; italics in original).

A key clusters' list of the DCC was created by utilising WordSmith Tools, before then comparing it to the BNC's Clusters' List (see Appendix 6.3). WordSmith Tools generated 138 key clusters for the DCC compared with the BNC, with Table 6.8 presenting the key clusters that occur in at least five of the DCC's texts.

Table 6.8 Key clusters of the DCC compared with the BNC

No.	Key clusters in the DCC	Freq.	Texts
1	morning noon and night	15	8
2	for old acquaintance sake	9	8
3	we understand each other	13	7
4	a sufficient reason for	9	7
5	a confused heap of	6	6
6	quite right said mr	6	6
7	by slow degrees and	6	6
8	by very slow degrees	6	6
9	pounds shillings and pence	11	5
10	the rain fell heavily	7	5
11	i'll stand by you	7	5
12	has seen better days	6	5
13	in exact proportion as	6	5
14	information relative to the	6	5
15	at arm's length	6	5
16	presented itself to the	5	5
17	the domestic economy of	5	5
18	sir roger de coverley	5	5
19	deeply sensible of the	5	5
20	be two parties to	5	5
21	brass plate on the	5	5
22	a favourable opportunity of	5	5
23	a gentle tap at	5	5
24	a miscellaneous collection of	5	5
25	an angry look at	5	5

26	low crowned broad brimmed	5	5
27	more shame for you	5	5
28	long ago when he	5	5

Considering the contexts of the key cluster *for old acquaintance sake* enables greater exploration of its usage as it occurs on nine occasions in eight texts. This key cluster is negative in comparison to the BNC, as it fails to occur in the reference of the key clusters extracted from the BNC. The word *acquaintance* suggests a marginal friendship, which may facilitate identification of the manner of the relationship between the characters with whom this key cluster is employed. Through examining the contexts of this cluster, it can be noted that it entails money in two incidents:

‘I’ll owe you five, Joe,’ said Arabella, ‘for old acquaintance sake, you know;’ and another most captivating smile was bestowed upon the corpulent intruder.

(The Pickwick Papers)

‘A shilling, dear!’ she said, with her eager avaricious face, ‘or sixpence! For old acquaintance sake. I’m so poor. And my handsome gal’-- looking over her shoulder--‘she’s my gal, Rob--half starves me.’

(Dombey and Son)

6.7 Application of grading the task of reading

In this section, I illustrate how those learners of English who aspire towards the ability to read Dickens’s works in their authentic text format can be assisted in this endeavour. As established in Chapter 2, the sophistication of Dickens’s texts necessitates a different approach to reading and comprehending his works. **The suggested approach is to grade the task of reading, rather than merely grading or simplifying the text itself, as applied to the simplified texts of Dickens or others’ works. Therefore, instead of simplifying the text, I intend to simplify the reading task itself by focusing primarily on introducing the lexical items within the target text and directing**

the learners gradually from part to part, for instance, while consolidating and building upon that which they have already acquired. Identifying those lexical items necessary to approach a given text is an essential step that equips learners with the required knowledge to commence proper reading and gain a sound understanding of the text.

Since the task of reading and comprehending authentic text represents a challenging one for non-native readers, regardless of whether they have been classified as advanced learners, any approach intended to scaffold learners in this objective should extend beyond merely suggesting that non-native learners read and re-read the texts to reveal the embedded meaning and thus achieve their goal of comprehension. In my approach, I intend to raise the learners' awareness in the domain of Dickens's lexicon and believe that tackling this challenge, as discussed previously in Chapter 2, can be successfully achieved through equipping the learners with the necessary vocabulary prior to reading the targeted texts. By introducing the vocabulary needed to engage with a specific work of Dickens I firstly address the learners' needs, prior to providing them with the reading task. Secondly, I simplify the reading task itself as opposed to the text, with Rixon (2000) suggesting that 'the teacher should grade the task rather than the text, so that a relatively impenetrable text can be given [as] an extremely easy task' (Rixon 2000: 68).

Beach et al. (2011) address the nature of literary language in claiming that '[e]ach historical era has its own characteristic features of linguistic style' (Beach et al. 2011: 65). Those cited features of linguistic style which can be associated with Dickens's works include 'long sentences, with multiple modifying phrases that are frequently embedded within other phrases and clauses', in addition to an 'uncharacteristic usage and occasional strange words' (Beach et al. 2011: 65). The level of lexical sophistication has been demonstrated in Chapter 2, since it represents the focus of this study.

Simultaneously, the actual sample passages that are analysed to demonstrate the lexical sophistication in Dickens's works also confirm the same fact regarding the length of his sentences. This characteristic of an extended sentence length is exemplified in Table 6.9, which features the first two sentences of *Oliver Twist* in their authentic and simplified forms. The *italics* in the authentic texts reflect the omitted content, which obviously results in a reduction in the extent of the detail and formation (the lexicon) conveyed in the simplified counterparts, together with the associated reduction in the complexity structure of said text.

Table 6.9 The first two sentences from *Oliver Twist*: authentic and simplified versions

Authentic version	Simplified version
<p>Among other public buildings in a certain town, <i>which for many reasons it will be prudent to refrain from mentioning, and to which I will assign no fictitious name</i>, there is one <i>anciently common to most towns, great or small: to wit</i>, a workhouse; and in this workhouse was born; <i>on a day and date which I need not trouble myself to repeat, inasmuch as it can be of no possible consequence to the reader, in this stage of the business at all events</i>; the item of mortality whose name is prefixed to the head of this chapter.</p>	<p>Oliver Twist was born in a workhouse, and when he arrived in this hard world, it was very doubtful whether he would live beyond the first three minutes.</p>
<p><i>For a long time after it was ushered into this world of sorrow and trouble, by the parish surgeon</i>, it remained a matter of considerable doubt whether the child would survive <i>to bear any name at all; in which case it is somewhat more than probable that these memoirs would never have appeared; or, if they had, that being comprised within a couple of pages, they would have possessed the inestimable merit of being the most</i></p>	<p>He lay on a hard little bed and struggled to start breathing.</p>

<i>concise and faithful specimen of biography, extant in the literature of any age or country.</i>	
--	--

It can be noted that Dickens utilises extended sentences that provide extensive details to draw vivid descriptions for the situation or characters. Through adopting such an approach, Dickens attempts to ‘exploit the resources offered by the linguistic system’ in order to produce texts with ‘syntactic complexity’ (Biber & Conrad 2009: 152).

6.8 Vocabulary sophistication

The rationale that lies behind studying Dickens’s lexicon in their co-text is to enable non-native learners to not only acquire the lexical items, but also to familiarise themselves with the manner in which such a lexicon is employed. Bowker (2012) indicates that exploring the lexicon contained within the authentic texts ‘can be a rich source of usage information’ that includes the ‘combinations of words’ and the ‘phraseology of a language’ (Bowker 2012: 387). Becoming acquainted with the manner in which language is employed scaffolds learners’ ability to develop their comprehension skills. Moreover, it assists learners in producing ‘more fluent’ language as a result of their being familiarised with the ‘fixed expressions and formulas’, as Singleton (2000: 55) suggests. Furthermore, Milton and Donzelli (2013) stress the significance of developing ‘a second language lexicon’ and how that can serve to support learners in coping with authentic texts (Milton & Donzelli 2013: 447).

Biber and Conrad (2009) consider the characteristics of ‘spelling and word choice’ as being the ‘most obvious difference between eighteenth- and twentieth-century novels’ (Biber & Conrad 2009: 151), with *Oliver Twist* offering evidence to support this assertion as demonstrated in Table 6.10, which features a number of the unconventional spellings contained within the text.

Table 6.10 Examples of unconventional spellings found in *Oliver Twist*

Unconventional spellings	Conventional counterparts
<i>acause</i>	because
<i>bisness</i>	business
<i>chimbley</i>	chimney
<i>gen'l'men</i>	gentleman
<i>hextricate</i>	extricate
<i>nothink</i>	nothing
<i>o'</i>	of
<i>obstinit</i>	obstinate
<i>sinds</i>	sends
<i>spectable</i>	spectacle
<i>sweepin'</i>	sweeping
<i>vereas</i>	whereas
<i>vith</i>	with
<i>wery</i>	very
<i>wot</i>	what

Regarding the examples of Dicken's lexical choice, a number of these examples will be discussed below in Section 6.10. Dickens's infrequent and unique vocabulary can be identified due to their low frequency, and then they can receive the focus they necessitate depending on when they occur in the target text.

In the following section, I generate the Dickens's Word List (DWL) by echoing those patterns employed in Nation's baseword lists, whereby the headwords are included with their families' members. I thus examine the DCC against Nation's baseword lists (described in Table 6.11) in order to reveal the details presented in Table 6.12.

Table 6.11 Paul Nation's baseword lists

File name	Number of headwords (groups/families)	Number of word types
1	1,000	6,857
2	1,000	6,370

3	1,000	5,880
4	1,000	4,865
5	1,000	4,294
6	1,000	4,102
7	1,000	3,679
8	1,000	3,419
9	1,000	3,196
10	1,000	2,982
11	1,000	2,942
12	1,000	2,754
13	1,000	2,415
14	1,000	2,299
15	1,000	2,283
16	1,000	2,086
17	1,000	2,076
18	1,000	1,933
19	1,000	1,872
20	1,000	1,820
21	1,000	1,651
22	1,000	1,539
23	1,000	1,394
24	1,000	1,296
25	1,000	1,675
26	1	1
27	1	1
28	1	1
29	1	1
30	1	1
31	21,662	22,409
32	38	196
33	3,108	6,044
34	1,083	1,149

Table 6.12 reveals the lexical profile statistics for the entire corpus of Dickens's works (the DCC), showing the number of tokens found in each list and their coverage in the entire corpus, the number of types and then the groups/headwords (lexemes) that were identified in each baseword list.

Table 6.12 The DCC's lexical profile statistics

Level	File	Tokens	Tokens%	Types	Groups
1	basewrd1.txt	5,081,983	81.93307%	4,863	991
2	basewrd2.txt	370,688	5.97633%	4,226	985
3	basewrd3.txt	163,244	2.631863%	3,393	933
4	basewrd4.txt	111,870	1.803598%	2,917	887
5	basewrd5.txt	69,780	1.125012%	2,425	844
6	basewrd6.txt	47,423	0.764566%	2,207	800
7	basewrd7.txt	31,434	0.506787%	1,812	723
8	basewrd8.txt	25,079	0.40433%	1,527	688
9	basewrd9.txt	18,815	0.30334%	1,441	668
10	basewrd10.txt	13,427	0.216474%	1,216	616
11	basewrd11.txt	10,671	0.172041%	953	532
12	basewrd12.txt	10,104	0.162899%	884	503
13	basewrd13.txt	7,497	0.120869%	655	412
14	basewrd14.txt	5,602	0.090317%	516	341
15	basewrd15.txt	3,291	0.053058%	423	288
16	basewrd16.txt	2,456	0.039596%	357	241
17	basewrd17.txt	3,923	0.063248%	304	214
18	basewrd18.txt	1,615	0.026037%	248	185
19	basewrd19.txt	1,472	0.023732%	183	144
20	basewrd20.txt	1,107	0.017847%	183	136
21	basewrd21.txt	785	0.012656%	138	108
22	basewrd22.txt	857	0.013817%	130	114
23	basewrd23.txt	495	0.007981%	129	117
24	basewrd24.txt	1,156	0.018637%	100	89
25	basewrd25.txt	939	0.015139%	85	80
31	basewrd31.txt	108,300	1.746041%	3,227	3,090
32	basewrd32.txt	15,073	0.243011%	60	28
33	basewrd33.txt	14,551	0.234595%	892	631
34	basewrd34.txt	3,153	0.050833%	147	144
0	Types not found in base lists	75,813	1.222277%	11,764	11,764
Total		6,202,603	100.00%	47,405	27,296

Table 6.12 presents 34 + 1 baseword lists arranged as *levels* (their frequency in the BNC and COCA). The *file* column indicates what is included in that list (i.e. the headwords and family members). As for the *token* and *type* columns, these show the number of tokens in the DCC that were identified in each level list, in addition to the number of word types. Lastly, the *group*

column simply refers to the number of headwords (basewords/roots/stem). The Dickens's Word List (see Appendix 6.4 for a sample) contains 48,333 words, which are categorised as follows: the headwords, the family members of each headword and the proper names (characters, places and things). After testing through the AntWordProfiler, it was found that this list provides 100% coverage of the entire corpus of Dickens's works, the DCC. Due to its substantial size and the impracticalities of including this entire list in the Appendices, the complete Dickens's Word List has been made available online at: <https://goo.gl/x6DmQ0>.

6.9 Guiding the reading of Dickens's works

The concept behind the Dickens's Word List can be employed to facilitate the accessibility of Dicken's works to non-native learners. I now demonstrate how such a list can be applied in the grading of the task of reading, again utilising *Oliver Twist* as an example of Dickens's work. Using AntConc 3.4.4w (Anthony 2005), *Oliver Twist* in the DCC consists of 165,925 tokens and 10,378 word types (see Table 6.13). Interestingly, the text only has 5,802 raw headwords, which I have manually revised to exclude in conventional spellings.

Table 6.13 Lexical profile statistics for *Oliver Twist*

Level	File	Tokens	Tokens%	Types	Groups
1	basewrd1.txt	135,745	81.81%	2,713	946
2	basewrd2.txt	9,768	5.89%	1,853	798
3	basewrd3.txt	4,428	2.67%	1,188	617
4	basewrd4.txt	3,050	1.84%	913	507
5	basewrd5.txt	2,036	1.23%	619	397
6	basewrd6.txt	1,311	0.79%	510	332
7	basewrd7.txt	798	0.48%	380	268
8	basewrd8.txt	702	0.42%	290	223
9	basewrd9.txt	510	0.31%	266	207
10	basewrd10.txt	368	0.22%	179	150
11	basewrd11.txt	654	0.39%	138	110

12	basewrd12.txt	258	0.16%	135	114
13	basewrd13.txt	112	0.07%	58	53
14	basewrd14.txt	126	0.08%	63	57
15	basewrd15.txt	55	0.03%	34	34
16	basewrd16.txt	108	0.07%	45	40
17	basewrd17.txt	173	0.1%	36	29
18	basewrd18.txt	34	0.02%	20	20
19	basewrd19.txt	32	0.02%	19	16
20	basewrd20.txt	13	0.01%	9	8
21	basewrd21.txt	14	0.01%	10	9
22	basewrd22.txt	6	0%	5	5
23	basewrd23.txt	18	0.01%	9	9
24	basewrd24.txt	12	0.01%	7	7
25	basewrd25.txt	6	0%	4	3
31	basewrd31.txt	3,426	2.06%	216	213
32	basewrd32.txt	402	0.24%	29	19
33	basewrd33.txt	395	0.24%	129	110
34	basewrd34.txt	95	0.06%	14	14
0	Not in any list	1,268	0.76%	487	487
Total		165,923	100.00%	10,378	(5,802) 5,707

Now, I categorise *Oliver Twist* into eleven files, referred to as 'parts' in the discussion below, with a view to maintaining a reasonable length of words between the files whilst also matching the possible length of reading task that undergraduate students may complete in a given study term ranging typically from twelve to fifteen weeks. Following the division of *Oliver Twist* into eleven files, each containing approximately five chapters as detailed in Table 6.14, I extract the number of headwords in each file (referred to as 'groups' in AntWordProfiler 1.4.0w). Then, assuming that the learners who read the first part of the novel (chapters 1–5) will not need to re-learn the same headwords again since they have already encountered them, I exclude the headwords previously occurring in the first part from the headwords in the second part. I repeat the same steps with the remaining nine files.

Table 6.14 Headwords in each part of *Oliver Twist*

Part	Chapters	Tokens	Headwords	Cumulative headwords	Percentage
1	01-05	19,504	2,383	2,383	41.76%
2	06-10	11,743	1,779	650	11.39%
3	11-15	15,570	1,944	526	9.22%
4	16-20	16,581	2,150	462	8.10%
5	21-25	11,912	1,881	314	5.50%
6	26-30	14,457	2,065	289	5.07%
7	31-35	17,368	2,024	270	4.74%
8	36-40	16,316	2,124	257	4.51%
9	41-45	15,026	1,970	208	3.65%
10	46-50	17,636	2,171	249	4.37%
11	51-53	9,810	1,583	99	1.74%
Total		165,926	22,074	5,707	100%

Table 6.14 demonstrates the journey that the readers of *Oliver Twist* travel in terms of vocabulary exposure. When reading the first five chapters, learners will encounter 19,504 words (tokens), of which only 2,383 are headwords. When they progress to the second part that covers chapters 6-10, they will read 11,743 tokens which contain 1,779 headwords; however, the readers will already have acquired a significant number of these, leaving them with only 650 new headwords to supplement those that they have previously learnt through inference by reading part one. What I mean by learning here is gaining the necessary knowledge and skills to acquire the targeted lexical items, a process which goes beyond merely reading the text once for the first time. This phenomenon is echoed for the remaining chapters, which explains the notion of the cumulative process of adding new words. The total number of headwords in *Oliver Twist* without repetition is 5,707, while the other column of headwords presents the number of headwords in each chapter and states the total number of these headwords regardless of their repetition from previous parts. All these lists produced from analysing

Oliver Twist and presented in Table 6.14 will be made available online, as detailed in Table 6.15.

Table 6.15 Lists resulting from the analysis of *Oliver Twist*

The novel divided into eleven parts collected in one zipped file	http://archive.org/details/OliverTwistIn11Parts
The headwords identified in each of the eleven parts	http://archive.org/details/HeadwordsOfEachPart
The cumulative headwords that each part adds to the one that precedes it	http://archive.org/details/CumulativeHeadwords

The same concept of grading the task of reading for learners of English, and particularly those readers of Dickens’s work, can be applied on a broader scale to include all fifteen of Dickens’s novels. This complete range appears chronologically according to Davis (1999), as presented in Table 6.16.

Table 6.16 Ranking Dickens’s fifteen novels chronologically

No.	Novel	Publishing date
1	The Pickwick Papers	1837
2	Oliver Twist	1838
3	Nicholas Nickleby	1839
4	The Old Curiosity Shop	1841
5	Barnaby Rudge	1841
6	Martin Chuzzlewit	1844
7	Dombey and Son	1848
8	David Copperfield	1850
9	Bleak House	1853
10	Hard Times	1854
11	Little Dorrit	1857
12	A Tale of Two Cities	1859
13	Great Expectations	1861
14	Our Mutual Friend	1865
15	The Mystery of Edwin Drood	1870

Through examining the number of headwords in each novel, I propose the sequencing of these works in a different order, namely one that is determined by the **lexical diversity**. **Lexical diversity refers in its basic concept to a measurement 'based on a comparison between the number of different words (types) and the total number of words (tokens)' (Malvern & Richards 2002: 87). Therefore, '[t]he degree to which new words are introduced and used in a text shows in the lexical diversity of the particular text. Lexical diversity can therefore be regarded as a measure for rich or varied language use' (Henrichs & Schoonen 2009: 5). By applying the measure of lexical diversity, Dickens's novels have been ordered whereby they progress from the less sophisticated novels in terms of their lexicon to their more advanced counterparts, as seen in Table 6.17.**

Table 6.17 Ranking Dickens's fifteen novels according to their headword diversity

Chronological order	Novel	Tokens	Vocabulary intensity (Headwords)
11	Little Dorrit	345,606	896
10	Bleak House	362,838	922
12	A Tale of Two Cities	138,366	928
13	Great Expectations	188,899	950
9	David Copperfield	363,477	973
8	Dombey and Son	363,605	1,034
14	Our Mutual Friend	333,798	1,083
7	Hard Times	105,663	1,152
6	Martin Chuzzlewit	346,740	1,328
5	Barnaby Rudge	260,427	1,552
4	The Old Curiosity Shop	222,511	1,815
15	The Mystery of Edwin Drood	97,850	1,874
3	Nicholas Nickleby	332,526	2,356
2	Oliver Twist	165,926	3,166
1	The Pickwick Papers	313,619	5,734

6.10 Dickens's lexical selections

The lexical selection represents one of the stylistic devices that impacts on the recipients, these being either the readers of a text or the listeners of a spoken account (see Shen 2014: 193; Hall 2014: 248; Fialho & Zyngier 2014: 333). Likewise, the lexical diversity (vocabulary richness) plays a role in the formation of a specific impression (Hosman 2008: 1122). In his analysis of the complex texture of several of John Milton's selected sonnets, Stockwell (2002) considers a number of the linguistic formal features that create the effects on readers, amongst which is the word choice as the 'poem's style' (Stockwell 2002: 85). Identifying the 'choice and patterns' in the language provides 'evidence of constitutive habits of representation' (Fahnestock 2009: 194), which enable enhanced understanding and appreciation of style.

Jeffries (2010) discusses the phenomenon of *naming and describing* as one of those linguistic resources that affect the function of the language. She addresses the question of 'how individual texts (and implicitly their authors) may choose from the regular resources of the language in representing a view of the world' (Jeffries 2010: 17). The manner in which an author, for instance, refers to something can result in a stylistic impact or an ideological effect, as Fairclough (2003) describes it, and either one is initiated by choosing a specific name or by modifying it in a specific manner. It is the same case when the events or actions are described without loaded verbs that are used to convey one type of message or another. Therefore, the wide range of lexical items that are found in Dickens's work can be investigated at differing levels, for instance by identifying the types of nouns, adjectives, verbs or adverbs which were selected to express an attitude or create a desired meaning. I will demonstrate through examples how the choice of a word can result in having a specific impact on the reader. The examples discussed below have been

intuitively selected from the alphabetically ordered Dickens's Lexical List, which explains why the examined words all begin with the letter *a*.

6.10.1 Analysis of word choice: *accoucheur*

The first word to be investigated is *accoucheur*, and in order to establish an idea about its frequency it is helpful to present these following points. Firstly, *accoucheur* occurs only once in the DCC: in *Great Expectations*, Chapter 4. The sentence in which this word appears reads as follows:

As to me, I think my sister must have had some general idea that I was a young offender whom an Accoucheur Policeman had taken up (on my birthday) and delivered over to her, to be dealt with according to the outraged majesty of the law.

It also occurs in the BNC on one occasion only in *The Possessed or, The Devils*; a novel in three parts by the Russian novelist Fyodor Dostoevsky, translated by Constance Garnett. The passage in which *accoucheur* occurs reads:

Our official town doctor Rozanov, himself an accoucheur, declared quite positively that on one occasion when a patient in labour was screaming and calling on the name of the Almighty, a free-thinking sally fired off like a pistol-shot by Mrs Virginsky struck such fear into the patient that delivery was greatly accelerated.

Other factors that indicate towards the rarity of utilising this word is its absence from some of the major English learner's dictionaries, and after consulting the *Oxford Advanced Learner's Dictionary*, *Cambridge Advanced Learner's Dictionary & Thesaurus*, *Cambridge Academic Content Dictionary*, *Longman Dictionary of Contemporary English* and *Collins COBUILD Advanced Learner's Dictionary*, I found that none of them list *accoucheur* as an entry. However, *accoucheur* is listed in the *Concise Oxford English Dictionary* and is defined as 'a male midwife'.

Due to the sparse concordance lines from which the learners can establish a pattern regarding the use of this word, learners must thus rely on the available sense of the lexical item, besides what the co-text may provide to help the readers view the reported event as accurately as possible. By considering the co-text in which *accoucheur* occurs in *Great Expectations*, the following observations can be made. Firstly, Pip was describing what he thought to be the perception of his sister (Mrs Joe Gargery) about him. In that sentence, Pip thinks that his 'sister must have had some general idea that [he] was a young offender' and that that young offender had been taken up by 'an Accoucheur Policeman' on his birthday. From this sentence, we learn that there is: *accoucheur policeman* / young offender. The adjective + noun *accoucheur policeman* being anaphoric for the *young offender*, as *accoucheur* refers to Pip's youth while *policeman* supports his status of being an alleged offender. Furthermore, *accoucheur* offers an indication of how early such a status was allocated, since he was effectively seen by an obstetrician during his birth. The use of *policeman* here can provide an additional impression, besides the fact that Pip is a young offender to be dealt with, which is the sense that the case is an emergency. However, is it an emergency due to the need to deal with a juvenile delinquent and young offender, or due to the necessity of assisting a lady in labour giving birth to a child who might then behave illegally in the future? There is no indication as yet that reveals which is more related to the use of that word, but both possibilities can be deemed acceptable in that situation.

6.10.2 Analysis of word choice: *round-aboutedly*

Another word to appraise is the compound adverb *round-aboutedly*, which does not appear in the BNC. I have also consulted *WebCorp Live*, which

provides access to the Internet as a corpus. This source returns only five instances, as follows:

- 1: the Nuns' House, was euphuistically, not to say round-aboutedly, denominated 'the apartment allotted to study,'
- 2: the Nuns' House, was euphuistically, not to say round-aboutedly, denominated 'the apartment allotted to study,'
- 3: So last week my mother informs me, sort of round-aboutedly, that the Evil Monkeys are invited to a Tea ...
- 4: the Nuns' House, was euphuistically, not to say round-aboutedly, denominated 'the apartment allotted to study,'
- 5: with unsolicited advice that will eventually and round-aboutedly get to that point: I recently came back from a ...

Three instances of these five (1, 2 and 4) are from Dickens's *The Mystery of Edwin Drood*, which is already included in the DCC, with the remaining two instances originating from personal blogs. By consulting the above-mentioned dictionaries for the words *round-about/round about*, but not *round-aboutedly*, only the *Cambridge Advanced Learner's Dictionary* has it as an entry and defines it as meaning '*approximately*'. I also referred to the *Concise Oxford Thesaurus*, where I found *round about* listed as having several synonyms: *approximately, about, around, circa, roughly, of the order of, something like, more or less, as near as dammit to, close to, near to, practically; or so, or thereabouts, give or take a few; not far off, nearly, almost, approaching; Brit. getting on for*. It can be noticed here that these synonyms are all adverbs, yet Dickens used *round-aboutedly* as an adverb only after adding the suffix *-edly*.

These two examples reveal the type of analysis that learners can be assisted to engage with once they have acquired the vocabulary necessary to understand Dickens's text in the first instance. The next stage will be to guide them through the process of understanding the effects of selecting specific lexical items over others, and the manner in which choice and usage in context can create the desired meanings and impact on the readers.

6.11 Exploring the DCC's key concepts

In this section, the DCC will be semantically explored through an investigation of key semantic concepts via Wmatrix, a web-based tool that can

be utilised for corpus analysis and comparison (Rayson 2002; Rayson 2008). As Rayson (2002) highlights, Wmatrix categorises the uploaded texts from a grammatical perspective into the respective parts of speech (POS). POS represents an annotation (morphosyntactic or tagging) which enables word-class labels (grammatical class) for the principle components of speech (e.g. nouns, verbs and prepositions), while also 'defining sub-classes, such as singular and plural nouns, positive, comparative and superlative adjectives, and so on' (Rayson 2002:20). Texts are processed by Wmatrix through the organisation and quantification of words, which Walker (2012) describes as follows:

- *Lexically (Word level)*: All the words in a text are ordered either alphabetically or by frequency of occurrence into frequency lists, with the most frequent words featuring first, and the frequency decreasing as the list descends.
- *Grammatically (Parts of speech)*: In this case every word from the corpus or uploaded text is attributed a tag that denotes the particular category of grammar or POS with which it is associated. The grammatical grouping of words is then employed to generate frequency lists, which are either ordered alphabetically, via the POS tag or frequency, where the POS tags of greatest frequency would be found at the top of the list.
- *Semantically (Semantic categories)*: Here all the words from the corpus or uploaded text are attributed semantic tags from a predefined group of 21 primary semantic fields (see Table 6.18). The semantic groupings are either listed alphabetically (by semantic tag) or by frequency (descending from the most frequent words) (see Walker 2012: 75–6).

In the interests of semantic analysis I uploaded the DCC corpus texts via a web browser, and automatically tagged the texts by either their POS or

semantically. Then, I compared the POS and semantic domain levels for their keyness, which resulted in the frequency profiles employed in the discussion below (see Payson 2008: 529). The use of key semantic concepts will enable a deeper assessment of the contents of the DCC in terms of the thematic topics that can be found with significant occurrence (to the point of overuse compared to a reference corpus) in Dickens's work. Table 6.18 presents the DCC's initial 20 *key* semantic concepts as generated by the Wmatrix tool, while harnessing the BNC Written Sampler as a reference corpus (see Appendix 6.5 for the complete table of all of the DCC's overused semantic concepts). The initial 20 keys from Appendix 6.5 can be found in Table 6.18 as follows:

Table 6.18 The DCC's first 20 key semantic concepts

No.	Semtag	Key semantic domains	DCCFreq%	BNCsampler%	LL
1	Z8	Pronouns	746,664 (13.26)	72,023 (7.44)	25760.47
2	B1	Anatomy and Physiology	70,609 (1.25)	5,489 (0.57)	4040.92
3	Q2.1	Speech: Communicative	70,083 (1.24)	7,024 (0.73)	2157.25
4	Z99	Unmatched	176,400 (3.13)	22,165 (2.29)	2102.64
5	A13.3	Degree: Boosters	51,919 (0.92)	4,808 (0.50)	2003.36
6	W2	Light	5,467 (0.1)	0 (0.00)	1735.05
7	M6	Location and Direction	85,596 (1.52)	9,859 (1.02)	1578.99
8	S2.2	People: Male	30,669 (0.54)	2,534 (0.26)	1550.88
9	A13	Degree	44,66 (0.08)	0 (0.00)	1417.37
10	Z6	Negative	70,158 (1.25)	8,052 (0.83)	1314.57
11	E4.1-	Sad	15,736 (0.28)	979 (0.10)	1296.38
12	T1.1	Time: General	4,888 (0.09)	47 (0.00)	1200.67
13	X3.4	Sensory: Sight	30,119 (0.53)	2,795 (0.29)	1155.83
14	M1	Moving, Coming and Going	82,979 (1.47)	10,157 (1.05)	1142.72
15	X2.2+	Knowledgeable	25,924 (0.46)	2,302 (0.24)	1112.44
16	E2+	Like	17,633 (0.31)	1,372 (0.14)	1007.36
17	L1+	Alive	4,835 (0.09)	93 (0.01)	968.76
18	E4.2+	Content	7,101 (0.13)	352 (0.04)	768.47
19	W2-	Darkness	2,212 (0.04)	0 (0.00)	702.02
20	M2	Putting, Pulling, Pushing, Transporting	44,174 (0.78)	5,347 (0.55)	644.59

The first column in Table 6.18 presents the DCC's first 20 key concepts as ordered by Log-Likelihood (LL). Meanwhile, the total number of significant semantic categories found in Appendix 6.5 is 172. The semantic tag employed by Wmatrix can be found in the second column; then followed by the overused key semantic concepts/domains in the third column. The DCCFreq%column presents the lexical items' frequencies in each semantic domain, while its percentage in the target/node corpus (DCC) can be found in parentheses. Next, the BNCSamply% column presents each semantic domain's lexical items in terms of the frequencies of occurrence in the reference corpus, again with its percentage included in parentheses. The final column presents the log-likelihood of the key semantic concepts. With a view to determining the key semantic domains of statistical significance, I applied the log-likelihood statistics suggested by Wmatrix in order to output solely those domains with a log-likelihood value greater than 6.63 as the cut-off point, to reflect 99% confidence in the results. It has been confirmed by Rayson (2008) that the result of the key comparison is that those key items of significance can be found near the head of the list as the result is organised on the basis of log-likelihood, which reveals the difference in terms of its significance. Moreover, in order to confirm the statistical significance, I will only examine items with a log-likelihood value over 6.63, and thus with 99% confidence of significance (Rayson 2008; Hu 2015).

Through appraisal of the initial category as an example, it can be noted that the first semantic category (semtag) is *Pronouns* (Z8). The lexical items related to this category occur in the DCC 746,664 times, representing 13.26% of Dickens's entire work (i.e. the DCC); meanwhile, the identical semantic category in the reference corpus features 72,023 lexical items, representing 7.44% of the BNC Written Sampler, which is one of several built-in reference corpora available for use in the Wmatrix tool. The log-likelihood for this

category was 25,760.47, which provides over 99% confidence of a statistically significant result, according to Rayson (2008). The Wmatrix tool presents plus (+) and (-) minus signs, with the (+) sign indicating towards the DCC's overused semantic concepts/categories in relation to the reference corpus, while the (-) sign implies underuse in the DCC in relation to the reference corpus. The lens of this section is solely focused on the exploration of the overused key concepts.

The significant key concepts found in Table 6.18 (see Appendix 6.5 for all the overused fields) comprise a wide range of overused semantic domains when measured against the reference corpus, the BNC Written Sampler. The complete USAS semantic tagsets presented in Table 6.19 highlight 21 important discourse fields (at the general level), which then extend into 232 subdivisions/structures that are multi-tiered in each category level (Archer, Wilson & Rayson 2002). Table 6.19 presents the 21 discourse fields as follows:

Table 6.19 The 21 discourse fields at the semantic hierarchy in general levels (Wmatrix)

A General & Abstract Terms	B The Body and the Individual	C Arts & Crafts	E Emotional Actions, States & Processes
F Food & Farming	G Government & Public Domain	H Architecture, Buildings, Houses & the Home	I Money & Commerce in Industry
K Entertainment, Sports & Games	L Life & Living Things	M Movement, Location, Travel & Transport	N Numbers & Measurement

O Substances, Materials, Objects & Equipment	P Education	Q Linguistic Actions, States & Processes: Communication	S Social Actions, States & Processes
T Time	W The World & Our Environment	X Psychological Actions, States & Processes	Y Science & Technology
Z Names & Grammatical Words			

I solely investigated those items categorised with a (+) code, with this symbol underscoring the semantic concepts as being overused in the DCC when compared to the BNC Written Sampler (see Table 6.18 and Appendix 6.5) and comprising of 172 categories in total. The top 20 items as measured by their log-likelihood values were included in Table 6.18, with their percentage of the respective corpora featuring in parentheses. Following a period of contemplation on the identified overused semantic fields, a number of these appear to have particular potential with respect to those concepts that frequently occur in Dickens's work. There is the potential, for example, to examine how Dickens presents *Anatomy and Physiology* as a thematic concept by exploring the word forms that are found in the B1 category, and considering the manner in which they describe the settings, scenes or characters' physical appearance. The total semantic categories, (e.g. there are 172 multi-tier structures/category labels in the DCC, out of a possible 232 semantic subdivisions) offers evidence that can be found in the key concepts presented in Table 6.18 (Appendix 6.5). The key concept of the greatest significance is the Z8 semantic domain (*Pronouns*), with 187 types and a total frequency of types in the DCC of 746,664 (13.26%). Nevertheless, since pronouns offer less value in respect to locating the topics or thematic concepts

that occur in the DCC, the Z8 semantic domain will not be discussed, outside of it featuring 187 word types.

6.11.1 The *Anatomy & Physiology* category (B1)

There are 879 word types shown in the B1 category (*Anatomy & Physiology*), with a total frequency of 70,609 types (tokens) representing 1.25% of the DCC, as calculated by Wmatrix. Table 6.20 presents the 100 most frequent lexical items in the DCC tagged in the B1 (*Anatomy & Physiology*) category, alongside their associated raw frequencies in the DCC.

Table 6.20 The 100 most frequent items in the *Anatomy & Physiology* category (B1) of the DCC

Number of types shown: 879
Total frequency of types (tokens) shown: 70,609 (1.25%) of the DCC
head (5,966), hand (5,761), face (5,208), eyes (4,919), hands (3,011), heart (2,752), arm (1,626), eye (1,560), hair (1,397), arms (1,396), back (1,275), feet (1,187), body (999), legs (997), lips (987), tears (950), foot (847), breast (797), mouth (758), sleep (735), nose (694), faces (687), shoulder (676), breath (662), heads (648), blood (607), ears (575), neck (564), ear (530), born (443), leg (433), fingers (424), forehead (403), chin (402), teeth (373), asleep (373), Fanny (372), finger (359), hearts (348), tired (334), bosom (334), throat (318), cheek (303), shoulders (293), roused (272), knees (268), tongue (265), unconscious (250), consciousness (246), knee (233), awake (229), slept (221), sleeping (210), chest (206), elbow (205), hearing (202), brow (189), bodies (185), eyebrows (168), awoke (163), limbs (162), birth (160), brain (159), cheeks (157), bones (156), breathing (155), waist (149), lip (148), waking (144), insensible (141), cell (140), go_to_bed (135), wake (124), organ (124), elbows (124), breathe (117), beard (117), bodily (116), went_to_bed (115), breathless (115), complexion (114), fist (113), crying (110), flesh (109), backs (109), thumb (108), tear (108), rouse (101), lap (101), breathed (100), fast_asleep (99), bald (98), slumber (95), frown (93), palm (92), physical (90), endurance (90), limb (87), drowsy (87), wrist (87)

In order to describe the features of anatomy and physiology, Dickens employs 879 lexical items, the 100 most frequent of which can be found in Table 6.20. Included in this category are body parts (e.g. *head, hand, face*),

bodily actions (e.g. *sleeping, panting, waking*), states of being (e.g. *asleep, tired, roused*), bodily conditions (e.g. *pregnant, tan, thirst*), and bodily products (e.g. *tears, spits, sweat*) (see Walker 2012). Nevertheless, there is value in noting that the Wmatrix lexicon is not 100% accurate, leading to some incidences of incorrect categorisation. For example, although the word *Fanny* was classified in B1 (*Anatomy & Physiology*), its actual use in the DCC is as a proper name, which can also be noted from the capitalisation. This name was employed to title a number of characters in the DCC: Fanny Brown in *Reprinted Pieces*, Fanny Squeers in *Nicholas Nickleby*, Mrs Fanny Dombey in *Dombey and Son*, Fanny Cleaver in *Our Mutual Friend*, Fanny Dorrit in *Little Dorrit* and Miss Fanny Wilson in *Miscellaneous Papers – Plays and Poems* (see Hawes 2002). The reason for its miscategorisation can be explained by referring to the literal meaning of the word *fanny*, which can refer to a part of the human anatomy. All the consulted dictionaries such as the CALD, OALD, COED, LDCE and the CCALD consider it to be ‘UK OFFENSIVE’, ‘Brit. vulgar slang’, ‘British English taboo informal a very offensive word ... Do not use this word’ and ‘BRIT INFORMAL, VERY RUDE’, respectively (capitals in original), since it refers to (1) the female sex organs (BrE, taboo, slang), and (2) a person’s bottom (slang, especially NAmE) (OALD). It is important for learners to be aware of this potential for error when using the corpus to explore aspects of Dickens’s work or other texts. It draws the learners’ attention to the fact that sharing the same orthographic conversion of spelling lexical items cannot be viewed in isolation from the co-text in which it occurs.

The reason for the B1 category (*Anatomy & Physiology*) being a key concept is due to Dickens’s physical descriptions of the characters in his works. Çakır (2011) highlights the fundamental nature of the character in fiction, where particular techniques are employed to create and extend the representation in the reader’s mind. Dickens can introduce and develop a

character through reporting, while conveying particular insights into the character's personality where there is significantly greater description of the characters' physical features and typically additional explanation, i.e. diegesis (telling) or elucidation through mimesis (showing) (Dennis 2007: 4; McIntyre & Walker 2011: 118). 'The diegetic level of a narrative is that of the main story, whereas the "higher" level at which the story is told is extradiegetic (i.e. standing outside the sphere of the main story)' (Baldick 2001: 66). Dickens works to create his characters by virtue of where they are seen to interact and transact, as opposed to merely offering a brief physical description. It is noted by Çakır (2011) that imaginative writers have a tendency to describe a character's appearance, whereas in fiction the character's physical appearance is often described in order to promote a 'mental picture or visualisation of the character' (Çakır 2011: 565–7). This description may 'include the physique and facial features, his clothes, his diseases, his bodily defects, his noticeable scars and warts. The author needs to use almost every particular active and unique trait about a character. Physical details are functional in the creation of a character' (Çakır 2011: 565–7).

Despite the frequency of occurrence of the first three words (*head* (5966), *hand* (5761), and *face* (5208) in category B1 being high, and putting to one side for a moment the fact that any investigation into every concordance line with its broader context would prove impractical from a time perspective, I will attempt to respond to the question of why these three words have the highest frequencies in this category. In the case of the word *head*, the physical appearance is a source of interest through which Dickens conveys a specific impression. This, as Balossi (2014) indicates, suggests the notion that the *head* is associated with the mind, where feelings and reason originate (Balossi 2014: 129). Moreover, Watkin (2009) observes that as a facet of his fiction Dickens's use of phrenology proposes that the shape of a character's *head* determines

particular personality traits (Watkin 2009: 181), e.g. 'like the head of an elephant in a state of melancholy madness' (*Hard Times*, Chap. 5). The *head* can also be a means by which Dickens presents a character's thoughts and emotions. As for the word *hand*, as it is used to convey several meanings, apart from the physical appearance of a character's gestures, to express the need to carry out an action, or to indicate towards independence, e.g. 'I posted with my own hand, and directed with my own hand' (*Dombey and Son*, Chap. LVI). With relation to anatomy and physiology, it can be noticed that the *hand* was described besides others as being 'right', 'immense', 'outstretched', 'smooth', 'white', 'bloodless', 'heavy', 'sunburnt', 'rough', 'small' and 'skinny', which again justifies having the word *hand* amongst the most frequent three words in this category. In the case of the word *face*, by analysing instances in which it is found it can be revealed that on the majority of occasions *face* refers to both the characters' physical attributes and their facial expressions. Awareness of the fact that Dickens's works feature more than 13,000 characters (Hawes 2002: ix), and whose faces may be described on occasion, would be a reasonable response to any uncertainty regarding the high usage of the word. Further explanation as to why we find Dickens employing references to anatomy and physiology is Dickens's tendency to focus externally in his description of his characters. Dickens avoids engaging with the feelings or emotional states of his characters, in preference for descriptions of what can be seen on their faces, such as their expressions.

Therefore, his natural *modus operandi* is to employ anatomical and physiological references, and on occasion to use these as metaphors to describe the emotions, with the *head*, *hand* and *face* being the most frequently referenced (see, for example, Andrews 2006, for clarity on how Dickens develops characters). The corpus approach adopted here not only confirms the critical positions regarding the manner in which characters are described,

but rather it further documents precisely how such lexical items related to body parts were overused and inserted by Dickens to serve different purposes in respect to his character creation and development. It would not be possible to manually identify the quantity (frequency) and the quality (usages as they appear in concordance lines) of such words without utilising the corpus approach. Çakır (2011) asserts that by placing the majority of the characters in his novels in differing situations, Dickens reflects on their personalities through their respective employment and action. Via their personas, Dickens's characters invariably expose their inner mechanisms and deeper personalities, without any explicit description of their inner lives. Sucksmith (1970) indicates that Dickens frequently describes the inner lives of his characters via 'the activity of the persona, a process which is not only psychologically accurate but typical of the way most people do express an inner life of which they are unconscious' (Sucksmith 1970, cited in Çakır 2011: 571). Action represents an effective and secure approach to introducing and extending characters, and is therefore a common strategy employed by Dickens, who 'describes their mannerisms, their reactions, and their behaviors towards other characters' (Çakır 2011: 570–1).

6.11.2 The *Unmatched* category Z99

It is presumed that the Z99 (*Unmatched*) category features those words employed by Dickens but that did not feature in the Wmatrix lexicon, and therefore were not able to be classified into any of the existing semantic fields. Presenting this category as a key semantic domain suggests that the DCC contains a higher number of *distinguished* words in comparison with the BNC Written Sampler. These unmatched words may represent, for example, lexical items that appear with less frequency such as *arfanarf*, less familiar proper nouns such as *Defarge*, or even those atypical spellings of other words such as

arter and *somethink*. Despite this category containing a total of 176,400 word types, some examples will be explored manually below to clarify certain issues that arise.

The reason for some words being classified into the Z99 (*Unmatched*) category is due to Dickens's approach to hyphenation, for example, *street-door* and *by-and-by*. There is also the case that some other words receive unconventional spelling; for example, the two words *gentleman* and *gentlemen* were spelt as follows:

genelman	genelmen
gen'leman	gen'lemen
genlman	genlemen
gen'l'man	gen'l'men
gent'lman	gent'lmen
genlmn	genlmns
gentlman	
genlemn	
genlm'n	
gen'lm'n	

Each one of the variants of *gentleman* and *gentlemen* were classified as *unmatched* in the Wmatrix lexicon, and were therefore placed into this category. This same phenomenon arises with other words with unconventional variants applied by Dickens in his writing. There is a third class of words that can be found in this category and which may be deemed uncommon; for example, (i) the names of people (e.g. *Volumnia*, *Noakes*, and *Plornish*), (ii) places (e.g. *Beauvais* and *Boulogne* in France and *Yarmouth*, an English town), or (iii) unique items (e.g. *bluchers*: historical strong leather half-boots or high shoes (COED); *chaise*: chiefly historical, a horse-drawn carriage, (COED); and *farthing*: an old British coin worth one quarter of an old penny (OALD)). Further examples include *almshouse*, *parlour* and *apothecary*. All of the above were placed in the Z99 category. Through investigation of the proper

nouns (excluding character names) or places identified in the DCC, this could illuminate aspects such as Dickens's geographical knowledge or preference for other places, and how he consequently integrated this into his writing. The above examples allude to non-native readers of Dickens in all likelihood requiring a broad cultural awareness in order to enable their appreciation of his works, especially where the use of such nouns is not arbitrary.

There is the potential here to form the observation that it may not be possible for this computer software to decode any meaning from those words that have been hyphenated together or in the case of proper nouns. A further point of observation is that Dickens demonstrates a high degree of flexibility in his creative use of the hyphen to introduce words or concepts into his writing, which can be ascribed to the freedom with which hyphens can be employed in the English language. Also found within the Z99 category are foreign words deriving from French and Italian. The French words include *château* (a large French country house or castle, OED), *claret* (red wine made in the region near Bordeaux in France) and *conciergerie* (the office or room of a concierge or porter's lodge; a state prison. During the Revolution it was the prison where the chief victims were confined prior to execution (DPF). Meanwhile, the Italian words include *altro* (other, different, CID), *festa* (day, holiday, festival or saint's day, CID), and *moccolo* (candle end, CID).

A further category within Z99 is archaic words. These words have been considered archaic i) due to the manner in which they have been classified by Wmatrix as *unmatched*, ii) since they have been classified as archaic in several of the glossaries of Dickens's published works (e.g. *A Tale of Two Cities* and *Great Expectations* by York Press), and iii) based on the learners' dictionaries that I have consulted, in addition to the supporting indication from the lexicon of the word processor (Microsoft Word). Examples of these archaic words include *a-bed* (in bed, COED), *doom* (judge's decision, sentence, COED),

hark (listen, CALD) and *yonder* (over there, CALD). Moreover, a number of colloquial words can be found in Z99 such as *codger* (old man), *jiggered* (euphemism for damned), and *squinting* (looking). It is possible to note from this sample of unusual word types that the vocabulary used by Dickens is wide, varied and contains a degree of uncommon items. The presence of these lexical items implies that on occasion Dickens may intentionally select unconventional words, as opposed to their more commonly used and perfectly functional counterparts.

This discussion regarding the *Unmatched Z99* category highlights that Dickens's lexical prowess is remarkable, and thus presents a particular challenge to non-native readers. In the sections below I consider the DCC's other key concepts, shifting the lens to focus on several of the significant categories. The key concept of Z99 (*Unmatched*), as a highly significant category in the DCC, has become surprisingly useful. Indeed, as the Wmatrix lexicon is employed as a comparison, it reveals that unusual nature of the DCC's lexicon. Moreover, empirical evidence has been provided for a number of the observations made regarding the diversity present in the vocabulary of Dickens. His prowess as a narrator clearly shines through via his broad knowledge and lexical diversity, which is clear from the countless facets that we encounter of Dickens's myriad characters since he describes them with such skill and vitality. The Z99 category can also be beneficial for learners of English by firstly allowing recognition of the variant spellings employed in a specific context and how they may represent other possible forms as a potential historical variant. Such variants can be categorised according to whether they are orthographical or phonological, such as representing a variety of spoken accent, for instance.

6.11.3 Light: key concepts in the DCC

There are a range of domains at the key-semantic level which are antonymous, as can be seen in the following examples: *B2- Disease vs B2+ Healthy*, *E3- Violent/Angry vs E3+ Calm* and *E4.1- Sad vs E4.1+ Happy*. Table 6.21 below presents all the antonymous key concepts in the DCC.

Table 6.21 Antonymous key-concepts in the DCC

No.	Semtag	Semantic domain	No.	Semtag	Semantic domain
68	A1.3-	No Caution	128	A1.3+	Cautious
49	B2-	Disease	111	B2+	Healthy
144	E2-	Dislike	16	E2+	Like
71	E3-	Violent/Angry	35	E3+	Calm
11	E4.1-	Sad	24	E4.1+	Happy
72	E5-	Fear/Shock	80	E5+	Bravery
127	E6-	Worry	77	E6+	Confident
98	G2.1-	Crime	117	G2.1+	Lawful
47	G2.2-	Unethical	26	G2.2+	Ethical
121	I1.2-	Debt-free	139	I1.2+	Spending and Money Loss
162	N3.3--	Distance: Near	160	N3.3+	Distance: Far
152	N3.7-	Short and Narrow	101	N3.7+	Long, Tall and Wide
79	N6-	Infrequent	34	N6+	Frequent
32	O4.2-	Judgement of Appearance: Ugly	38	O4.2+	Judgement of Appearance: Beautiful
119	O4.6-	Temperature: Cold	129	O4.6+	Temperature: Hot/On Fire
3	Q2.1	Speech: Communicative	113	Q2.1-	Speech: Not Communicating
124	S1.1.4-	Undeserving	61	S1.1.4+	Deserving
76	S1.2.3-	Unselfish	97	S1.2.3+	Selfish
118	S1.2.4-	Impolite	27	S1.2.4+	Polite
89	S1.2.5-	Weak	154	S1.2.5+	Tough/Strong
31	S2.1	People: Female	8	S2.2	People: Male
60	S7.2-	No Respect	29	S7.2+	Respected
6	W2	Light	19	W2-	Darkness
28	X2.1	Thought, Belief	40	X2.2-	No Knowledge
94	X2.5-	Not Understanding	171	X2.5+	Understanding
22	X2.6-	Unexpected	66	X2.6+	Expected
39	X3.2-	Sound: Quiet	44	X3.2+	Sound: Loud
168	X3.4-	Unseen	83	X3.4+	Seen
84	X5.1-	Inattentive	140	X5.1+	Attentive

64	X5.2-	Uninterested/Bored /Unenergetic	130	X5.2+	Interested/Excited/Energetic
74	X9.2-	Failure	166	X9.2+++	Success

Two antonymous key concepts are explored below to enable demonstration of how harnessing the semantic analysis feature of Wmatrix can allow greater understanding of Dickens's usage, while illuminating the manner in which he utilises lexical items in order to introduce thematic concepts. *Light* and *Darkness* are the two antonymous key-semantic concepts selected, since I find them to be the two most prevalent antonymous key concepts in the DCC. As presented in Table 6.18, row 6, the semtag W2 *Light* occurs with a frequency of 5,467 in the DCC, representing 0.10% of the corpus, while its relative frequency and percentage of the BNC Sampler Written are 0 and 0.00%, respectively, highlighting the extent of this key concept's over-utilisation when compared to the BNC Sampler, providing a log-likelihood of 1735.05. Meanwhile, the semtag W2- *Darkness*, found on row 19, has a frequency of 2,212, representing 0.04% of the DCC, and again its relative frequency and percentage of the BNC Written sampler are 0 and 0.00%, respectively, providing a log-likelihood of 702.02.

Table 6.22 Key semantic concepts of the *Light* (W2) category in the DCC

Number of types shown: 82

Total frequency of types shown: 5,467 (0.10% of the DCC)

light (2,509), lights (302), lighted (275), shone (243), shining (231), moonlight (132), daylight (123), shine (111), sunshine (101), ray (97), gleam (96), lightning (91), lighting (91), beams (88), sunlight (87), rays (75), lighter (64), gleaming (63), shines (60), lustre (53), lightest (53), guppy (50), gleamed (42), beam (39), lightness (35), brilliancy (28), starlight (24), lighthouse (24), illuminated (24), illumination (23), glimmer (18), firelight (14), candle-light (14), rainbow (13), candlelight (12), gleams (11), glimmering (10), torchlight (9), rainbows (9), glimmered (9), streamers (8), moonshine (7), moonlit (7), lighthouses (7), sunbeams (6), sunbeam (6), illumined (6), illuminate (6), fire-light (6), lightsome (5), night-light (4), streamer (3), lamp-light (3), illuminations (3), glow-worms (3), Bradbury (3), sun-beam (2), moon-light (2), lightnings (2), lamplight (2), BRADBURY (2), torch-light (1), scintillation. (1),

phosphorescent (1), moonless (1), moon-lit (1), lustres (1), look_light (1), lightnesses (1), lighthouse-keeper (1), light_style (1), light. (1), light-footed (1), light- (1), lamplight. (1), illumining (1), illumines (1), illumine (1), illuminates (1), gup-py (1), glow-worm (1), --light (1)

To demonstrate the manner in which Dickens generally employs the semantic concept *Light*, I now consider the most frequently occurring word in this category: 'light'. Since it would be both complicated and impractical to investigate every single one of the 2,509 occurrences of this word in this semantic category, after consulting with AntConc 3.4.4w I find that the first word which collocates with 'light' in the DCC is 'eyes'. Table 6.23 reveals how the word 'light' collocates with the word 'eyes' in the DCC.

The rationale behind the selection of the most frequently occurring word in this category for investigation is, as Sinclair (1991) suggests, that collocates can provide 'a semantic analysis of a word' (Sinclair 1991: 115–6). Hunston (2002) also indicates to the fact that 'patterns of association', namely 'how lexical items tend to co-occur, are built up over large amounts of text and are often unavailable to intuition or conscious awareness. They can therefore convey messages implicitly and even be at odds with an overt statement' (Hunston 2002: 109). The word 'light' is found to collocate with 698 items, limiting the minimum frequency of co-occurrence to five times. To allow identification of the first content word to collocate with 'light', it is essential to exclude the function words, and thus I use AntConc 3.4.4w for this purpose. Following the exclusion of these function words, which while offering little significance are typically the most frequent collocates of any particular word, 'eyes' appears to be the first content word to collocate with 'light'. Through consideration of the usage of the item 'light' alongside its collocate 'eyes', it can be found that Dickens has a tendency to utilise the semantic concept of light in order to develop a range of creative meanings. The potential senses of the words 'light' and 'eyes' as individual lexical items

are not pertinent to this discussion, since these have been well documented in learners' dictionaries. The objective here is to attempt to shift the lens of focus to the concept, image and particular meaning or sense created by Dickens in relation to the use of both 'light' and 'eyes', since they are treated as collocates in a range of contexts throughout the DCC. The result of the search for the content word collocates with 'light' in the DCC produced 545 collocates. The ten most frequently occurring collocational items are presented in Table 6.23 below.

Table 6.23 The ten most frequently occurring collocates for the word 'light'

Rank	Freq	Freq (L)	Freq (R)	Stats	Collocate
1	81	26	56	4.5305	day
2	80	56	25	3.55996	little
3	79	26	54	5.99914	fire
4	75	30	45	8.033	candle
5	68	30	45	2.25658	said
6	66	45	23	4.18171	made
7	61	39	27	3.90161	like
8	54	29	32	4.59709	eyes
9	54	27	27	4.29321	face
10	54	27	27	5.28506	light

From these ten collocates, only two items are found to be related to the B1 semantic concept of *Anatomy & Physiology*, as discussed above. Next, I investigated how the first item that relates to semtag B1 (*Anatomy & Physiology*), i.e. 'eyes', collocates with the first item that relates to the semtag W2 *Light*, i.e. 'light'. A search of the collocates between 'light' and 'eyes' resulted in the sixty-six occurrences presented in Table 6.24, alongside their respective concordance lines.

Table 6.24 How 'light' collocates with 'eye' in the DCC

1	along them, with reverently shaded	eyes, a bridge of light appeared
2	our boat, and, as my	eyes adapted themselves to the light
3	is performed!' She raised her	eyes again; and the light of
4	bright, curling hair, their sparkling	eyes, and their beautiful light tread,
5	to be ghostly in the	light, and the eyes of the
6	this one with the blue	eyes and light hair. This is
7	impart new lustre to his	eyes, and to light up new
8	up a hand between his	eyes and the light, and the
9	of Barbara in the full	light and lustre of his eyes--
10	court? They were not such	eyes as shun the light; they
11	until, when she turned her	eyes away, bars of light would
12	kiss to the child, whose	eyes beamed with a strange light
13	Two more, and a blue-	light burnt. All eyes watch the
14	obeyed, and Defarge followed the	light closely with his eyes. "Stop!--
15	Wilding to himself, as his	eyes enthusiastically followed the light
16	hair shaken down, her dark	eyes flashing with a raging light,
17	his lips, and a strange	light flashing in his filmy eyes.
18	in a window, and her	eyes, full of light, were fixed
19	certainly,' said Mr. Pickwick, his	eyes glistening with delight at this
20	go in together. As the	light goes in, the great eyes
21	posthumous child. My father's	eyes had closed upon the light
22	weak legs and very weak	light hair. One of his eyes
23	that little boy with the	light hair and pink eyes, in
24	a weak voice, white teeth,	light hair, and surprised eyes, some
25	sizes too small, when our	eyes happened to alight on a
26	rejoined the old man. 'My	eyes have seen more light than
27	she hurried on; the same	light in her eyes, the same
28	he retorted, with a wild	light in his eyes. "You do?"
29	upon her cheek, the bright	light in her eyes, and the
30	Mrs Dombey to understand '--the	light in his eyes fell upon
31	unassailable position?' Again the	light in his eyes fell upon
32	world!" said Richard with a	light in his eyes. My husband
33	that, I saw a boyish	light in his eyes that looked
34	for there was a cunning	light in his eyes as he
35	time to time with twinkling	eyes, in the light of a
36	said Bella, busily. 'Hold the	light, John. Shut your eyes, sir,
37	still is growing, and such	eyes light the waves as they
38	of me. I feared the	light might dazzle her eyes and
39	put his hand over his	eyes, murmuring that the light was
40	that he stood in the	light of the dark, bright eyes.
41	and went out of the	light of the dark, bright eyes,
42	his manner, You are the	light of my eyes. You are
43	tone of her voice, the	light of her eyes, the interest,
44	me I had been the	light of his eyes--indeed the

45	her joy and pride, the	light of her eyes, the happy
46	of her voice, in the	light of her eyes, in the
47	from the dust, that the	light of the eyes of the
48	crime; he saw the dawning	light of intelligence in the eyes
49	before, was the saddened softened	light of the once proud eyes;
50	and beauty's dress; the	light of women's eyes, outshining
51	hair, and in its thoughtful	eyes reflecting back their light, became
52	blind and curtain before sleepers'	eyes, shed light even into dreams,
53	and twisted into knots; his	eyes shone with a terrible light;
54	Mr. Pickwick's	eyes sparkled with delight, as he
55	you, though?' returned Kit, his	eyes sparkling with delight. 'Where are
56	his face flushed and his	eyes sparkling with delight, heedless of
57	the dark lashes of her	eyes sullenly veiling their light, that
58	The visitor, more with his	eyes than by the slight impulsive
59	music to him, and the	eyes that had been light, fell
60	hold so dear, as the	eyes that light our pleasures here.
61	to shut out from her	eyes the little light there was,
62	little, and will save your	eyes, too, from the light. My
63	accord, they all turned their	eyes towards the light of the
64	mended his pace, keeping his	eyes upon the light with a
65	conceivable, in her brother's	eyes, with as slight an affirmative
66	me. You see that red	light yonder?' They raised their eyes,

Through manual checking and reference to the wider context I investigated how 'light' collocates with 'eyes' in the DCC, with the following points to be noted. The manner in which 'light' and 'eyes' are employed in the DCC allows *five* possible consistent senses. The *first sense* concerns the most common and literal sense of the word 'light' as referring to the brightness that emerges from the sun, fire, lamps, etc., and which allows things to be seen (CALD). For instance, line 1 in Table 6.24 conveys this sense of 'light' as a physical brightness emitted from the sun, and something to celebrate due to its strength and vitality.

But, the glorious sun, rising, seemed to strike those words, that burden of the night, straight and warm to his heart in its long bright rays. And looking along them, with reverently shaded eyes, a bridge of light appeared to span the air between him and the sun, while the river sparkled under it.

(*A Tale of Two Cities*, Chap. IX)

Through examination of this excerpt, however, it can be noted that the senses and feelings aroused by the collocation of 'light' and 'eyes' extend beyond mere acknowledgement and appreciation of the bright rays of the sun, by forging an association between 'the glorious sun', the 'bridge of light' and Sydney's warm heart that lead to his decision to carry out an admirable act, in keeping with Dickens invariably utilising 'light' and 'eyes' in positive contexts in his works. The corpus stylistics approach offers insight through identification of the association between the characterisation and the mentioning of body parts. Mahlberg (2013) reveals how findings that emerge from analysing clusters can highlight the word building components of text world theory and the textual characterisation triggers.

The *second sense* relates to meeting by chance or coming across (Roget's II), where line 25 in Table 6.24 reads 'when our eyes happened to alight on a few suits of clothes ranged outside a shop-window' (*Sketches by Boz*, Chap. VI). The association here between 'light' and 'eyes' is defined by the use of the two-part verb 'alight on', which itself has etymologically originated from the word 'light' (see OED, LDCE). Furthermore, the use of 'happened to' also supports this interpretation of a chance meeting or coming across.

The *third sense* is in the description of the endearing features of a character through depicting the 'eyes' as an essential component of the character's beauty, thus enabling the reader to form a vivid image regarding their appearance. The context of line 6 reads as follows:

This is the lad, sir; this one with the blue eyes and light hair. This is a swimmer, sir, this fellow--a diver, Lord save us! This is a boy, sir, who had a fancy for plunging into eighteen feet of water, with his clothes on, and bringing up a blind man's dog, who was being drowned by the weight of his chain and collar, while his master stood wringing his hands upon the bank, bewailing the loss of his guide and friend.

(*The Old Curiosity Shop*, Chap. 52)

The same sense can be found with line 9, which refers to Barbara's admirable and attractive qualities (*The Old Curiosity Shop*, Chap. 69).

The *fourth sense* that is created through the collocation of 'light' with 'eyes' is the identification of a cause or source of great pleasure. Line 59 exemplifies this sense, which reads in the wider context as:

A young girl--his little grand-daughter--was hanging about him, endeavouring, with a thousand childish devices, to engage his attention; but the old man neither saw nor heard her. The voice that had been music to him, and the eyes that had been light, fell coldly on his senses. His limbs were shaking with disease, and the palsy had fastened on his mind.

(*Pickwick Papers*, Chap. XLII)

Despite the word 'delight' (Table 6.24, lines 19, 54–56) not originating from the word 'light', but rather it being of Middle English origin from the Old French 'lilitier' or 'to charm' (OED), 'eyes' were employed to symbolise the cause or source of significant pleasure: 'the eyes that had been light' – the eyes of the little girl – represent the grandfather's source of light and happiness.

A further example can be found on line 54: 'Mr. Pickwick's eyes sparkled with delight, as he sat and gloated over the treasure he had discovered' (*Pickwick Papers*, Chap. IX). The 'eyes', and the brightness with which they appear, convey the character's feelings of great joy and delight at his discovery of a perceived treasure, that is, an inscribed stone which he believes has significant archaeological value: he 'presumed the stone to be ancient, but solemnly denied the antiquity of the inscription' (*Pickwick Papers*, Chap. IX).

The *fifth sense* is antonymous to the previous, fourth sense and appears on line 53 from *Oliver Twist, Fagin's Last Night Alive*, where it reads:

His red hair hung down upon his bloodless face; his beard was torn, and twisted into knots; his eyes shone with a terrible light; his unwashed flesh crackled with the fever that burnt him up.

(*Oliver Twist*; Chap. LII)

Here, Dickens employs both 'light' and 'eyes' to convey the fear and rage that Fagan felt before his execution, allowing the light emitting from his eyes to qualify the inner turbulence of the character, and his feelings of impotency at his impending demise.

6.11.4 *Darkness*: key concepts in the DCC

The other key semantic concept in the DCC is *W2- Darkness*, representing a key concept that is antonymous from the previously discussed *W2 Light*. Table 6.25 presents the lexical items identified by Wmatrix as representing the key semantic concept *W2-Darkness* in the DCC.

Table 6.25 Key semantic concepts of the *Darkness* (W2) category in the DCC

Number of types shown: 10

Total frequency of types shown: 2,212 (0.04%)

dark (1,714), darkness (408), darkly (78), unlighted (4), glowering (2), darkneses (2), unilluminated, (1), on_the_dark_side (1), looked_dark (1), darkling (1)

In keeping with the above analysis, I examined the most frequently occurring item in this category, namely the word 'dark'. Moreover, to reduce the number of concordance lines where the word 'dark' has been included, I employed the same strategy of excluding the function word collocates in order to ensure a pertinent series of results. There are a total of 533 collocates of 'dark' in the DCC. Following the exclusion of the 166 function word

collocates, as conducted previously with the semantic concept *Light*, 367 collocates remain. Table 6.26 presents the DCC's 10 most frequently occurring content words that collocate with the word 'dark', out of the 367 identified.

Table 6.26 The 10 most frequent collocates with 'dark' in the DCC

Rank	Freq	Freq (L)	Freq (R)	Stats	Collocate
1	112	36	76	5.62752	night
2	103	9	94	5.89633	eyes
3	86	26	60	5.60638	room
4	82	11	71	7.36813	hair
5	62	30	32	3.64275	upon
6	59	23	36	3.80781	man
7	56	28	28	6.42449	dark
8	55	19	36	4.18206	like
9	55	8	47	6.74466	corner
10	55	35	20	3.54495	little

It is interesting to note that 'eyes', from the *B1 Anatomy & Physiology* category, is present among the most frequent content word collocates, as will be discussed in greater detail below. Rather, 'eyes' is the most frequent collocate with 'dark', which belongs to the *Anatomy & Physiology* (B1) category, again putting itself forward for investigation. Therefore, to extend our understanding Table 6.27 presents the concordance lines of 'dark' and its 99 collocates with 'eyes' from the DCC.

Table 6.27 'Dark' and its 99 collocates in the DCC

1	pair— 'Oh Powers of Heaven! what	dark eyes meets she there? 'Tis—'tis
2	surface, and beckoned him to approach;	dark gleaming eyes peered from the water
3	over his arm, and her beautiful	dark eyes fixed themselves upon his face
4	in a sweet, soft pair of	dark eyes, without feeling queer? I can'
5	Jew uttered these words, his bright	dark eyes, which had been staring vacantly
6	at the boy with his large	dark eyes. 'Who would have thought it!
7	by that alone. His face is	dark, like his hair and eyes; and,
8	They were tall stately figures, with	dark flashing eyes and hair of jet;
9	laudatory remarks touching his beautiful	dark eyes, and his sweet smile, and
10	lovely, bewitching, and not nineteen.	Dark eyes, long eyelashes, ripe and ruddy
11	of words; old Arthur Gride and	dark eyes and eyelashes, and lips that

12 could see that she raised her dark eyes to the face of her
13 locksmith's rosy daughter, before whose dark eyes even beef grew insignificant, and
14 man he had encountered in the dark last night. His eyes met those
15 of night, you may see his eyes in my dark room, shining like
16 table; 'ME,' he repeated, bending the dark hollows of his eyes upon him
17 said her father, putting back the dark hair from her sparkling eyes, 'to
18 gentlemen who were easily moved by dark eyes; and efforts were made to
19 chocolate-coloured face and his bright dark eyes and white teeth, came creeping
20 bed, and, fixing on her two dark eyes whose brightness was exaggerated
21 the deep wonder of her bright dark eyes, 'abroad.' 'Abroad, Martin!' 'Only
22 , but dirty yellow. He had bright dark eyes, which he kept half closed;
23 I do!' Mr Tapley opened his eyes wide in the dark; but did
24 in, before their depth was fathomed. Dark eyes, that reflected back the eyes
25 defined for an old man, and dark, bright, penetrating eyes, looked round
26 : 'Father, I am lonely in the dark. I want my eyes, my patient,
27 the direction of the speaker's eyes, she saw a dark figure standing
28 happy!' mused the other, bending his dark eyes upon the stooping figure, with
29 still almost strangers. Sissy, with her dark eyes wonderingly directed to Louisa's
30 good voice, good breeding, well-dressed, dark hair, bold eyes.' All which Mrs.
31 were young and hansom. Wi' fine dark thinkin eyes, and a still way,
32 never left behind; plain to the dark eyes of her mind, as the
33 evening sky, were plain to the dark eyes of her body; Mrs. Sparsit
34 creeping things that be. With her dark eyes and her hook nose warily
35 steady glare in her great round dark eyes that embarrassed me. Said my
36 her perfectly colourless face and deep dark eyes towards him; but without loosening
37 the thoughtful hours. Why did the dark eyes turn so often from this
38 , she was a little child. With dark eyes and hair, I recollect, and
39 dressed and very handsome, whose dark proud eyes were fixed upon the
40 her dark hair shaken down, her dark eyes flashing with a raging light,
41 agitated, quieted down; while her dark eyes, fixed upon the fire, exchanged
42 looked towards him, in the obscure dark corner, her speaking eyes, more earnest
43 angry at his intrusion, and the dark lashes of her eyes sullenly veiling
44 him in her beauty, with the dark eyes that had never turned away
45 not?' Edith signified assent with her dark eyes. 'Why?' returned Florence
46 all else, she sat with her dark eyes cast down, waiting for someone.
47 once. She stood there, with her dark disdainful eyes again upon him; and
48 say so?' Edith sat with her dark eyes gazing steadfastly before her, and
49 and there were tears in her dark eyes. 'I trust myself to that,'
50 with no shape at all, and eyes so dark that they seemed to
51 cruel relish. Again, I see her dark eyes roll round the church when
52 me as could come from his dark eyes. 'It rankled in your baby
53 From the first moment of her dark eyes resting on me, I saw
54 out and leaves him in the dark, with the gaunt eyes in the
55 made, and good looking, with crisp dark hair, bright eyes, and a broad
56 --some young lady, perhaps." His bright dark eyes glanced at me for the
57 look with the fullness of his dark eyes and broad forehead. "Humph!

58	in its face and curious little	dark marks under its eyes like faint
59	Mademoiselle Hortense fixes her black	eyes upon him with a dark frown
60	to a handsome girl with lustrous	dark hair and eyes, and very neatly
61	on, when she lifted up her	dark eyes and made this unexpected.
62	Meagles, shaking his head at the	dark eyes with a quiet caution. 'Take
63	maid. She had large unfeeling handsome	eyes, and dark unfeeling handsome hair,
64	certain, and is she fair or	dark blue eyes or black I wonder,
65	young woman, Mr Clennam, with very	dark hair and very dark eyes. If
66	spot. It flashed out of her	dark eyes as they regarded him, quivered
67	way. They were bright, handsome,	dark eyes,--what are called laughing eyes
68	than encounter those two handsome,	dark, bright eyes. One day, when Private
69	stood in the light of the	dark, bright eyes. There was a silent
70	out of the light of the	dark, bright eyes, an altered man. In
71	wherever Captain Taunton, with the	dark, bright eyes, led, there, close to
72	two friends, Major Taunton, with the	dark, bright eyes, and Ensign Richard
73	For God's sake!" The bright,	dark eyes--so very, very dark now,
74	pass the window; and the bright,	dark eyes of his debased time seemed
75	so old as that her bright,	dark eyes were dimmed--and remembering
76	one handsome elderly lady, with very	dark eyes and gray hair, that I
77	with her gray hair and her	dark eyes--hide her daughter's child
78	round my face, and that my	dark blue eyes were considered expressive,
79	quite complete, from the sparkling	dark eyes under his knowing uniform cap
80	of a person groping in the	dark. Long after his eyes had opened,
81	bright little comely girl with large	dark eyes. Looking full at me, the
82	short dark hair. He was a	dark man altogether, with good eyes and
83	village below, where it is already	dark, when I raise my eyes, and
84	five feet nine; black hair; complexion	dark; generally, rather handsome visage; eyes
85	rather handsome visage, complexion	dark, eyes dark, thin, long and sallow
86	had overtaken him when it was	dark night. He rubbed his eyes and
87	been suspected (perhaps was, by the	dark eyes that looked contemptuously at
88	Lucie had occupied. Madame Defarge's	dark eyes followed her through this rapid
89	waxwork and skeleton seemed to have	dark eyes that moved and looked at
90	in detail, his large head, his	dark complexion, his deep-set eyes, his
91	for the first time with his	dark deep-set eyes, 'we must revert
92	and the fog was heavy and	dark. Animate London, with smarting eyes
93	the lovely little face with her	dark, fiery eyes, and tenderly caressing the
94	gleam of fire in the intense	dark eyes, though they were then softened
95	against myself, Helena?' She, whose	dark eyes were watching the effect of
96	-haired person of thirty, with big	dark eyes that wholly wanted lustre, and
97	Tartar who saved him.' Helena's	dark eyes looked very earnestly at the
98	step was still ready, and his	dark eyes brightened at every happy
99	faces beam, sweet lips smile, and	dark eyes gleam; All these charms have

Through reference to the broader contexts of the uses of 'dark' together with its collocate 'eyes', I identified that two key senses are revealed. The *first*

sense refers to the eyes and their physical appearance, where two observations are made: (i) this is primarily the case when Dickens is describing the female character, and (ii) this collocation is used to complement their beauty.

Therefore, it is used positively in this sense. The *second sense* refers to the description of a character's feelings and attitudes, where two observations are again made: (i) male characters are commonly described instead of female (as in the previous sense), and (ii) the collocation is used in the sense of being bitter or unkind, and thus is deployed negatively. Lines 3 and 4 exemplify the first sense of a complementary description of a female in the following context:

I mention the circumstance, to show what a very uncommon sort of person this beautiful young lady must have been, to have affected my uncle in the way she did; he used to say, that as her long dark hair trailed over his arm, and her beautiful dark eyes fixed themselves upon his face when she recovered, he felt so strange and nervous that his legs trembled beneath him. But who can look in a sweet, soft pair of dark eyes, without feeling queer? I can't, gentlemen. I am afraid to look at some eyes I know, and that's the truth of it.

(*Pickwick Papers*, Chap. XLIX)

In other lines such as 6, 8 and 11 the same sense is again repeated in respect to the description of physical appearance.

Regarding the second sense, which describes the feelings and attitude of a typically male character, line 5 exemplifies this and can be found in the original content as follows:

As the Jew uttered these words, his bright dark eyes, which had been staring vacantly before him, fell on Oliver's face; the boy's eyes were fixed on his in mute curiosity; and although the recognition was only for an instant--for the briefest space of time that can possibly be conceived--it was enough to show the old man that he had been observed.

(*Oliver Twist*, Chap. IX)

6.12 Conclusion

Cortes (2002: 131) cites Bieber et al.'s (1999) study, where a large corpus was digitally analysed in order to identify those lexical bundles that appeared with frequency in academic and conversational genres, with complete lists of those lexical bundles that are categorised grammatically emerging as a result. Cortes (2002) then proceeds to investigate a corpus featuring written texts produced by first year university students in order to identify the most frequently utilised four word bundles and then compare these with Bieber et al.'s (1999) corpus, while additionally analysing the bundles' grammatical and functional characteristics.

When considering the vast trove of **semi-fixed expressions** and their potential benefit to learners of second and foreign languages, **since** the acquisition of additional **phrases and expressions of a language** is synonymous with language learning, **the value of introducing phraseology into teaching is reported by Granger (2011), together with how it can be distilled to provide fluency, which has a higher pedagogical priority. Lexical phrases can thus be employed by learners to aid in their spoken fluency; however, the inclusion of lexical phrases in classroom tuition can only be justified when their role in SLA and production is better understood.** Boers (2009) points towards the inevitability of the need to assign a priority of importance in such learning contexts, where there is invariably limited available time with which to study the language, and thus the pertinence of being able to define appropriate criteria for determining which lexical phrases can be justified in respect to their inclusion in the curricula of the language learning classroom.

In the context of using corpus in the classroom, Littlemore (2009) underscores the potential of employing language corpora as a 'way of introducing learners to all the senses at once' (Littlemore 2009: 55), with the

strategy entailing the offering to the language learner of several cases of the target language that have been extracted from a corpus containing authentic language use, and then requesting that they proceed to analyse and develop possible meanings for the language item under consideration, that can then be inserted into category diagrams they create themselves. Littlemore (2009) attests to the merits of such discovery tasks, and that where the categories of meaning accurately reflect the sense of the words under consideration, then the storage and recall of such metalinguistic characteristics can be more meaningful and have greater longevity than those scenarios where the categories of meaning are merely provided by the tutor.

In asserting the benefits of language corpora for the language learning classroom Littlemore (1999) believes the aforementioned discovery tasks have the potential to accelerate understanding of associated categories of meaning though facilitating learners' ability to engage with the full range of meaning scenarios linked with a particular word, thus promoting flexibility in understanding and the skills of shifting between differing senses of meaning, while also encouraging further research to investigate whether the study of language learning through this approach can lead to acquired knowledge.

The value of Dickens's corpus to language learners who have expressed an interest in accessing and understanding his discourse is that it allows the learners to gain access and exposure to the real and authentic language of Dickens himself (as opposed to the translated meaning sourced from dictionaries) 'instead of having to rely on intuitions and make-up examples' (Schmitt 2000: 68).

Fox (2010) **mentions** additional more generalised approaches that employ corpora to gain inferential understanding, a number of which are reliant on the analysis of semantics that can be considered to be both

traditional and formal, whereby a semantic analysis of the target texts is realised, together with any hypothesis that is required to be tested. This typically necessitates ‘a broad-coverage deep syntactic analysis, comprehensive semantic analysis, and a robust theorem prover’, while in the case of particular objectives such as responding to questions, the matching of semantic patterns can suffice (Fox 2010: 426). Further information sources may be required for analysis in order to establish the concordance between the information comprised within a text or set of texts, and the hypothesis that is to be tested (Fox 2010).

The qualitative analysis conducted in Section 6.11 discussed some of the most frequently occurring key semantic concepts – the *Anatomy and Physiology (B1)* category and the *Unmatched (Z99)* category – before proceeding to examine the two antonymous key semantic concepts of *Light (W2)* and *Darkness (W2-)*, selected since they represent the DCC’s most frequent antonymous key semantic concepts. Next, these most frequent items were investigated with regards to their most frequent collocates, with the possible senses that can be inferred from such combinations demonstrated. Interestingly, the most frequent semantic concept was *Anatomy and Physiology (B1)*. Then, the choice was guided by the most frequent antonymous categories: *Light (W2)* and *Dark (W2-)*. That the collocational item of the highest frequency in the *Anatomy & Physiology* semantic concept for both ‘light’ and ‘dark’ was the same lexical item – ‘eyes’ – is of great interest. It is hoped that this analysis serves to reveal the potential for enhanced learning regarding Dickens’s style, in respect to his choice of lexis and the manner in which meaning can be coded through the medium of collocation and associated context.

Chapter 7 Conclusion

Studying texts can be achieved through conducting analyses of language usage at multiple levels by highlighting those linguistic features that structure the text's form and contribute towards the creation of its meanings. Stylistic analysis allows the text to be approached from these different levels (i.e. semantic, syntactic, phonological, etc.), with each of these tiers having its own particular features. The stylistic analysis of texts is dependent upon the linguistic characteristics and facets of the language that can be explored on the basis of the measurable features of the texts; especially in terms of their being of quantitative and statistical interest (e.g. see [McIntyre 2008](#); [Mahlberg 2009](#); [Mahlberg & McIntyre 2011](#)). The linguistic study of literary texts, as one of the examples of discourse type, establishes the aim of arrival at an objective interpretation. Therefore, stylistics would contribute no novel approach to the interpretations of such texts unless this were to be achieved by addressing the texts through an empirical and objective methodology, and thus realising objective findings. In this context, the use of statistics is indispensable to the field of stylistics as it provides an important systematic key that leads to an enhanced awareness and understanding of the detailed linguistic features of a given text by, for example, noting and exploring the literary phenomenon. The statistical process contributes towards rationalising the texts' interpretation, language teaching and literary criticism by eliciting, providing and linking the interpretations to that evidence which is inherited from the text. This describes the basis on which this study has been constructed, and which is manifested in the corpus stylistic approach.

Stylistic studies apply statistics and measurement to assess, for instance, any deviation from the norm of the language's usage or the frequency of some patterns in the language in general, and primarily in

literary texts. The quantity itself is a factor of the visibility, since it draws the reader's attention to particular features in the texts resulting from those linguistic characteristics being positioned so unusually that the prospective language user can capture this characteristic usage due to its frequency. Deviation in the language in general, or in literary texts in particular, is identified by quantitative measurement, for example, through being significantly frequent or infrequent; otherwise, it would be deemed as an idiosyncrasy arising from the general patterns of the language in use. The language in this sense becomes statistically measurable, which lends the interest in stylistic studies towards identifying the linguistic characteristics, before then analysing and interpreting them after the generation of statistical data regarding the frequency and recurrence of the respective linguistic features. This exploration of the language is conducted by harnessing those tools available in the field of linguistics which can, for instance, facilitate in determining literary phenomenon. Thus, stylistics transforms the quantity profile of the linguistic features into a more accurate, objective interpretation of the texts based on the foundations of linguistically measurable characteristics, as opposed to the adopting of perspectives grounded in personal and subjective judgment.

Regarding the works of Charles Dickens, the above-described methodology was selected to study, firstly the selection that Dickens makes for his choice in lexicon, and furthermore how he employs it within different contexts. The manner in which Dickens exercises his selection in terms of lexical items is revealing of particular insight into his stylistic ability. One of the advantages of such selection is that it leads to the language, which is dependent upon the employment of linguistic features which are to all intents and purposes limitless in terms of creating meaning and producing a variety of styles, as shown in Section 6.10, or even defining the writers' styles

themselves, due to the various relations that can be identified between the linguistic levels in a given discourse.

Lexical items acquire additional senses through their interaction with the other language features available within the context of the text. Notwithstanding their referral to a particular primary meaning, lexical items reveal that they are simultaneously imbued with other possible meanings that are triggered by the contexts in which they occur. Therefore, the consideration of Dickens's lexicon (or lexicons, if each text were to be analysed individually) can be achieved by investigating the semantic fields in his works through a statistical methodology which is to be found in the corpus stylistic approach. Corpus stylistics contributes towards identifying the effective role of frequencies that constitute Dickens's own unique semanticity, and the manner in which such details surrounding the behaviour of these lexical items can be exploited in order to comprehend and teach his works.

One of the achievements of this study is in the construction of the Charles Dickens Complete Corpus (**the** DCC), which comprises comprehensively all of the works of Dickens according to the National Edition (a set of 40 volumes, the last two volumes of which are dedicated to Dickens's life by his appointed biographer, John Forster). At the current point in time of conducting and presenting this research, I am unaware of the existence of any other complete corpus of Dickens's works, as discussed in Chapter 4. It is hoped that this thesis's contribution to the research community will lend itself towards the more detailed and systematic investigation of the works of Dickens, while enabling attempts to explore Dickens as a linguistic phenomenon himself through responding to the questions of why his books have never failed to attract new readers since their first publication, while neither have his works since struggled to engage publishers in the reproduction of his literature. Dickens's writings have

exceeded the time and the geography within which he was located, to become widely known to students of literature from across the globe, and to English language learners virtually everywhere. The contribution of the DCC lies in providing complete materials that can be readily studied and analysed, so as to enable greater exploration of those linguistic facets that are inherent in Dickens's writings, which make his work unforgettable, and which provide a rationale for why his works continue to be read with such interest. The creation of the DCC should allow researchers the means with which to identify scientific and objective answers to those questions that they may have in relation to Dickens's works, thus enriching those linguistic studies centred on the author's works, and offering the distinctive pattern of stylistic and linguistic studies that flow between novelty and systematicity in the exploration of his works.

This study also achieves another of its main objectives, namely to facilitate the works of Dickens to non-native speakers in terms of identifying the lexical items employed through the corpus-based approach in order to analyse his entire body of texts, whilst focusing on the semantic level of his language (i.e. Dickens's lexicon, as in his wordlists, keywords and collocations). Furthermore, the study has produced evidence-based results that directly contribute towards the empowerment of the learner's focus on the lexical items as an initial step through which to scaffold greater comprehension of the text, as the study confirms a well-established relationship that exists between the knowledge of vocabulary and comprehension of the lexicon that is employed.

The results include the production of the DCC Headwords List for the complete works of Dickens (See Appendix 6.5, which due to its size has been made available online only at <https://archive.org/details/TheHeadwordsOfDCC7.1>); these can be exploited

as a foundation for a curriculum focused on developing understanding of those lexical items utilised by Dickens, as they are particularly associated with the works of Dickens rather than guested samples. This study has also been able to establish a compensative list of the DCC's Family Words; that is, the headwords plus their family members which fully span the entire works of Dickens. As it contains 102,753 words, and is thus too substantial to be included as an appendix within this study, I make this comprehensive list available for further investigation online at <https://archive.org/details/DCCBasewordList.Lemmatised.AlphabeticalOrder>. This resource can be employed in learning materials similar to the work of Diane Schmitt and Norbert Schmitt's (2005) *Focus on Vocabulary: Mastering the Academic Word List*, which is based on the Academic Word List produced by Averil Coxhead (2000). The analysis of Dickens's lexical selection also demonstrates how learners can be assisted to enhance their appreciation of Dickens's style in terms of the lexicon and semantic level employed in his work. Such analysis will enable learners to explore Dickens's patterns in respect to his selection and usage of lexical items, besides allowing learners to compare such usages with other reference corpora in order to establish more rigorous information regarding the behaviour of a given lexical item.

Finally, the DCC can be engaged with to conduct further research besides stylistic and literary studies, and that are related to lexicography and lexical studies, grammatical studies, register variation and genre analysis, dialect distinction and language variety, translation studies, diachronic study and language change, and language learning and teaching, as has been discussed in this thesis. Finally, additional related studies might consider addressing subjects concerned with semantics and discourse analysis. [Informing the commentary of literary critics with established linguistic evidence can be achieved through engaging with the approach of corpus](#)

stylistics. Therefore, literary criticisms of Dickens's works can be considered through testing the observations of literary critics while adopting a corpus-driven approach, whereby certain pre-existing claims or observations require examination in order to establish whether they are supported by the empirical evidence emerging from Dickens's corpus, the DCC.

Bibliography

- Abrams, Meyer H. 1999. *A glossary of literary terms*. 7th edn. Fort Worth: Harcourt Brace College Publishers.
- Adolphs, Svenja. 2006. *Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies*. Oxon: Routledge.
- Adolphs, Svenja & Ronald Carter. 2002. Corpus stylistics: point of view and semantic prosodies in *To The Lighthouse*. *Poetica* 58. 7–20.
- Aghagolzadeh, Ferdows & Farzaneh Tajabadi. 2012. A Debate on Literature as a Teaching Material in FLT. *Journal of Language Teaching and Research* 3(1). 205–210.
- Aluísio, Sandra M., Lucia Specia, Thiago A. S. Pardo, Erick G. Maziero, Helena M. Caseli & Renata P. M. Fortes. 2008. A corpus analysis of simple account texts and the proposal of simplification strategies: First steps towards text simplification systems. *Proceedings of the 26th annual ACM international conference on design of communication*, 15–22. Lisbon, Portugal: ACM.
- Andrews, Malcolm. 2006. Performing character. In John Bowen & Robert L. Patten (eds.), *Palgrave advances in Charles Dickens studies*, 69–89. Basingstoke: Palgrave Macmillan.
- Anthony, Laurence. 2005. AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. *International Professional Communication Conference Proceedings. IEEE 2005*, 729–737. Limerick Conf. & Sports Park Limerick, Ireland: IEEE Advancing Technology for Humanity.
- Anthony, Laurence. 2013. *AntWordProfiler 1.4.0w [Computer Software]*. Tokyo, Japan. Waseda University. Available from: <http://www.laurenceanthony.net> (Accessed: 14 October, 2015).
- Archer, Dawn, Andrew Wilson & Paul Rayson. 2002. Introduction to the USAS category system. <http://ucrel.lancs.ac.uk/usas/usas%20guide.pdf> (27 July, 2016).
- Argamon, Shlomo, Kevin Burns & Shlomo Dubnov (eds.). 2010. *The structure of style: Algorithmic approaches to understanding manner and meaning*. Heidelberg: Springer.
- Auerbach, Elsa R. & Denise Burgess. 1985. The hidden curriculum of survival ESL. *TESOL Quarterly* 19(3). 475–495.
- Baldick, Chris. 2001. *The concise Oxford dictionary of literary terms*. 2nd edn. Oxford: Oxford University Press.
- Bagarić, Vesna & Jelena M. Djigunović. 2007. Defining communicative competence. *Metodika Review paper* 8(1). 94–103.
- Baker, Paul. 2006a. *Using corpora in discourse analysis*. London: Continuum.

- Baker, Paul. 2006b. The British English 2006 Corpus. Available from: <http://www.ling.lancs.ac.uk/profiles/paul-baker> (Accessed: 25 March, 2015).
- Baker, Paul, Andrew Hardie & Tony McEnery. 2006. *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
- Balossi, Giuseppina. 2014. *A corpus linguistic approach to literary language and characterization: Virginia Woolf's the waves*. Amsterdam: John Benjamins Publishing Company.
- Bateman, John. 2008. *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Basingstoke: Palgrave Macmillan.
- Bauer, Laurie & Paul Nation. 1993. Word families. *International Journal of Lexicography* 6(4). 253–279.
- Beach, Richard, Deborah Appleman, Susan Hynds & Jeffrey Wilhelm. 2011. *Teaching literature to adolescents*. 2nd edn. Oxon: Routledge.
- Beard, Adrian. 2003. *The language of literature*. London: Routledge.
- Behnam, Biok. 1996. A stylistic study of cohesive features in English prose fiction with some pedagogical implications for non-native contexts. Sheffield: Department of English Language and Linguistics, University of Sheffield. PhD thesis.
- Bennett, Gena R. 2010. *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan: University of Michigan Press.
- Berg, William J. & Laurey K. Martin-Berg. 2002. A stylistic approach to foreign language acquisition and literary analysis. In Holly Tucker & Virginia Mitchell Scott (eds.), *SLA and the literature classroom: Fostering dialogues*, 173–191. Boston: Heinle & Heinle.
- Bernhardt, Elizabeth B. 2011. *Understanding advanced second-language reading*. London: Routledge.
- Bex, Tony, Michael Burke & Peter Stockwell (eds.). 2000. *Contextualized stylistics: In honour of Peter Verdonk*. Amsterdam: Rodopi.
- Biber, Douglas. 2011. Corpus linguistics and the study of literature: Back to the future? *Scientific Study of Literature* 1(1). 15–23.
- Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, Douglas, Edward Finegan, Stig Johansson, Susan Conrad & Geoffrey N. Leech. 1999. *Longman grammar of spoken and written English*. 1st edn. Harlow: Longman.
- Birch, David. 1989. *Language, literature, and critical practice: ways of analysing text*. London: Routledge.

- Bowker, Lynne. 2012. Meeting the needs of translators in the age of e-lexicography: Exploring the possibilities. In Sylviane Granger & Magali Paquot (eds.), *Electronic lexicography*, 379–398. Oxford: Oxford University Press.
- Bowker, Lynne & Jennifer Pearson. 2002. *Working with specialized language: A practical guide to using corpora*. London: Routledge.
- Bradford, Richard. 1997. *Stylistics*. London: Routledge.
- Breen, Michael P. 1985. Authenticity in the Language Classroom. *Applied Linguistics* 6(1). 60–70. doi:10.1093/applin/6.1.60.
- Browne, M. Neil & Stuart M. Keeley. 2007. *Asking the right questions: A guide to critical thinking*. 8th edn. Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Brown, Keith & Jim Miller. 2013. *The Cambridge dictionary of linguistics*. Cambridge: Cambridge University Press.
- Brumfit, Christopher & Ronald Carter. 1986. *Literature and language teaching*. Oxford: Oxford University Press.
- Brussino, Gabriella & Cathy Gunn. 2008. Australasian language learners and Italian web sites: A profitable learning partnership? In Felicia Zhang & Beth Barber (eds.), *Handbook of research on computer-enhanced language acquisition and learning*, 1–19. Hershey, PA: Information Science Reference.
- Burke, Michael. 2010. *Literary reading, cognition and emotion: An exploration of the oceanic mind*. London: Routledge.
- Burkhanov, Igor. 2003. Pragmatic specifications: Usage indications, labels, examples; dictionaries of style, dictionaries of collocations. In Piet van Sterkenburg (ed.), *A Practical Guide to Lexicography*, 102–113. Amsterdam: John Benjamins Publishing Company.
- Busse, Beatrix & Dan McIntyre. 2010. Language, Literature and Stylistics. In Dan McIntyre & Beatrix Busse (eds.), *Language and Style*. Basingstoke: Palgrave Macmillan.
- Bussmann, Hadumod. 1996. *Routledge dictionary of language and linguistics*. (Trans.) Gregory Trauth & Kerstin Kazzazi. London: Routledge.
- Butler, Paul. 2008. *Out of style: Reanimating stylistic study in composition and rhetoric*. Logan: Utah State University Press.
- Çakır, Hasan. 2011. *Dickens and his memorable characterization*. 565–573. Sarajevo, Bosnia and Herzegovina.
- Carlisle, Joanne F. 2007. Fostering morphological processing, vocabulary development, and reading comprehension. In Richard K. Wagner, Andrea E. Muse & Kendra R. Tannenbaum (eds.), *Vocabulary acquisition: Implications for reading comprehension*, 78–103. London: The Guilford Press.

- Carrell, Patricia L. 1984. Schema Theory and ESL Reading: Classroom Implications and Applications. *The Modern Language Journal* 68(4). 332–343.
- Carrell, Patricia L. 1991. Second language reading: Reading ability or language proficiency? *Applied Linguistics* 12(2). 159–179.
- Carter, Ronald. 1986. A question of interpretation: An overview of some recent developments in stylistics. In Theo D’haen (ed.), *Linguistics and the study of literature*, 7–26. Amsterdam: Rodopi.
- Carter, Ronald. 1996. Look both ways before crossing: Developments in the language and literature classroom. In Ronald Carter & John McRae (eds.), *Language, literature and the learner: Creative classroom practice*, 1–15. London: Longman.
- Carter, Ronald. 1997. *Investigating English discourse: language, literacy and literature*. London: Routledge.
- Carter, Ronald. 1998. *Vocabulary: Applied linguistic perspectives*. 2nd edn. London: Routledge.
- Carter, Ronald. 2010. Methodologies for stylistic analysis: practice and pedagogies. In Dan McIntyre & Beatrix Busse (eds.), *Language and Style*, 55–70. Basingstoke: Palgrave Macmillan.
- Carter, Ronald. 2012. Coda: Some rubber bullet points. *Language and Literature* 21(1). 106–114.
- Carter, Ronald & Michael N. Long. 1991. *Teaching literature*. London: Longman.
- Carter, Ronald & John McRae. 1996. *Language, literature and the learner: Creative classroom practice*. London: Routledge.
- Carter, Ronald & Paul Simpson (eds.). 1989a. *Language, discourse and literature: An introductory reader in discourse stylistics*. London: Routledge.
- Carter, Ronald & Paul Simpson. 1989b. Introduction. In Ronald Carter & Paul Simpson (eds.), *Language, discourse and literature: An introductory reader in discourse stylistics*, 1–20. London: Routledge.
- Cook, Guy. 1992. *The discourse of advertising*. London: Routledge.
- Cook, Guy. 1994. *Discourse and literature: The interplay of form and mind*. Oxford: Oxford University Press.
- Cook, Guy. 2008. Hocus pocus or God’s truth: the dual identity of Michael Stubbs. In Andrea Gerbig & Oliver Mason (eds.), *Language, people, numbers corpus linguistics and society*, 305–334. Amsterdam: Rodopi.
- Coxhead, Averil. 2000. A new academic word list. *TESOL Quarterly* 34(2). 213–238.
- Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

- Crossley, Scott A., David Allen & Danielle S. McNamara. 2012. Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research* 16(1). 89–108.
- Crossley, Scott A., Max M. Louwrese, Philip M. McCarthy & Danielle S. McNamara. 2007. A linguistic analysis of simplified and authentic texts. *The Modern Language Journal* 91(1). 15–30.
- Crossley, Scott A., Hae Sung Yang & Danielle S. McNamara. 2014. What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language* 26(1). 92–113.
- Cruse, Alan. 2006. *A glossary of semantics and pragmatics*. Edinburgh: Edinburgh University Press.
- Crystal, David & Derek Davy. 1969. *Investigating English style*. London: Longman.
- Culpeper, Jonathan. 2002. Computers, language and characterisation: An analysis of six characters in Romeo and Juliet. In Ulla Melander-Marttala, Carin Ostman & Merja Kyto (eds.), *Conversation in Life and in Literature: Papers from the ASLA Symposium*, vol. 15, 11–30. Uppsala: Association Suedoise de Linguistique Appliquee.
- Culpeper, Jonathan. 2009. Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics* 14(1). 29–59.
- Daller, Helmut, James Milton & Jeanine Treffers-Daller. 2007. Conventions, terminology and an overview of the book. In Helmut Daller, James Milton & Eanine Treffers-Daller (eds.), *Modelling and assessing vocabulary knowledge*, 1–32. Cambridge: Cambridge University Press.
- Davidson, T. T. L. 1990. Teaching with the Oxford Concordance Program. *Literary and Linguistic Computing* 5(1). 81–85.
- Davies, Alan & Henry G. Widdowson. 1974. Reading and writing. In John P. B. Allen & Stephen P. Corder (eds.), *Techniques in applied linguistics*, 155–201. Oxford: Oxford University Press.
- Davis, Paul B. 1999. *Charles Dickens A to Z: the essential reference to his life and work*. New York: Checkmark Books.
- Dennis, Helen May. 2007. *Native American literature: towards a spatialized reading*. London: Routledge.
- Emmott, Catherine. 1997. *Narrative comprehension: A discourse perspective*. Oxford: Oxford University Press.
- Fahnestock, Jeanne. 2009. Quid pro nobis. Rhetorical stylistics for argument analysis. In Frans H. van Eemeren (ed.), *Examining argumentation in context: Fifteen studies on strategic maneuvering*, 191–220. Amsterdam: John Benjamins Publishing Company.
- Fairclough, Norman. 2003. *Analysing discourse: Textual analysis for social Research*. London: Routledge.

- Fialho, Olivia & Sonia Zyngier. 2014. Quantitative methodological approaches to stylistics. In Michael Burke (ed.), *The Routledge handbook of stylistics*, 329–345. (Routledge Handbooks in English Language Studies). London: Routledge.
- Fidler, Masako Ueda. 2003. Reading and technology in less commonly taught languages and cultures. In Peter C. Patrikis (ed.), *Reading between the lines: Perspectives on foreign language literacy*, 99–117. New Haven: Yale University Press.
- Firth, John Rupert. 1951. Modes of meaning. *Essays and Studies (The English Association)*. 118–149.
- Fischer-Starcke, Bettina. 2009. Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis. *International Journal of Corpus Linguistics* 14(4). 492–523.
- Fischer-Starcke, Bettina. 2010. *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*. London: Continuum.
- Fish, Stanley. 1973. What is stylistics and why are they saying such terrible things about it? *Approaches to poetics*. 109–152.
- Flowerdew, John. 2013. *Discourse in English language education*. London: Routledge.
- Fowler, Roger. 1981. *Literature as social discourse: the practice of linguistic criticism*. Bloomington: Indiana University Press.
- Fowler, Roger. 1996. *Linguistic criticism*. Oxford: Oxford University Press.
- Gardner, Dee. 2007. Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics* 28(2). 241–265.
- Gascoigne, Carolyn. 2008. Independent second language reading as an interdependent process. In Stella Hurd & Tim Lewis (eds.), *Language learning strategies in independent settings*, 67–83. Bristol: Multilingual Matters.
- Gass, Susan M. 1988. Second language vocabulary acquisition. *Annual Review of Applied Linguistics* 9. 92–106.
- Gass, Susan M. 2003. Input and Interaction. In Catherine Doughty & Michael H. Long (eds.), *The handbook of second language acquisition*, 224–255. Oxford: Blackwell Publishing.
- Gass, Susan M. & Larry Selinker. 2008. *Second language acquisition: An introductory course*. 3rd edn. London: Routledge.
- Gavioli, Laura. 2005. *Exploring corpora for ESP learning*. Amsterdam: John Benjamins Publishing.
- Gee, James P. 1999. *An introduction to discourse analysis: Theory and method*. London: Routledge.
- Gilquin, Gaëtanelle. 2010. *Corpus, cognition and causative constructions*. Amsterdam: John Benjamins Publishing Company.

- Gilroy, Marie & Brian Parkinson. 1996. Teaching literature in a foreign language. *Language Teaching* 29(04). 213–225.
- Goulden, Robin, I. S. Paul Nation & John Read. 1990. How large can a receptive vocabulary be? *Applied Linguistics* 11(4). 341–363.
- Granger, Sylviane (ed.). 1998. *Learner English on computer*. London: Longman.
- Granger, Sylviane. 2011. From phraseology to pedagogy: Challenges and prospects. In T. Herbst, P. Uhrig & S. Schüller (eds.), *Chunks in Corpus Linguistics and Cognitive Linguistics: A tribute to John Sinclair*, 123–146.
- Green, Keith. 2000. The tip of the iceberg: Real texts, long texts and mental representations. In Tony Bex, Michael Burke & Peter Stockwell (eds.), *Contextualized stylistics: In honour of Peter Verdonk*, 49–66. Amsterdam: Rodopi.
- Gries, Stefan T. 2009a. *Quantitative corpus linguistics with R: a practical introduction*. London: Routledge.
- Gries, Stefan T. 2009b. What is Corpus Linguistics? *Language and Linguistics Compass* 3(5). 1225–1241.
- Hall, Geoff. 2005. *Literature in Language Education*. Basingstoke: Palgrave Macmillan.
- Hall, Geoff. 2007. Stylistics in Second Language Contexts: A Critical Perspective. In Greg Watson & Sonia Zyngier (eds.), *Literature and stylistics for language learners: theory and practice*, 91–105. Basingstoke: Palgrave Macmillan.
- Hall, Geoff. 2014. Pedagogical stylistics. In Michael Burke (ed.), *The Routledge handbook of stylistics*, 239–252. (Routledge Handbooks in English Language Studies). London: Routledge.
- Halliday, Michael A. K. 1971. Linguistic function and literary style: an inquiry into the language of William Golding's *The Inheritors*. In Seymour Chatman (ed.), *Literary style: A symposium*, 330–365. Oxford: Oxford University Press.
- Halliday, Michael A. K. 1991. Corpus studies and probabilistic grammar. In Karin Aijmer & Bengt Altenberg (eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, 30–40. London: Routledge.
- Halliday, Michael A. K. 1999. The notion of "context" in language education M. A. K. Halliday. In Mohsen Ghadessy (ed.), *Text and context in functional linguistics*, 1–24. Amsterdam: John Benjamins Publishing Company.
- Halliday, Michael A. K. 2007. *Language and education*. (Ed.) Jonathan J. Webster. London: Continuum.
- Halliday, Michael A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Hamilton, Craig. 2008. Applied rhetoric and stylistics in Anglo-Saxon countries in the 20th and 21st centuries. In Ulla Fix, Andreas Gardt &

- Joachim Knape (eds.), *Rhetoric and stylistics: An international handbook of historical and systematic research*, vol. 1, 550–569. Berlin: Walter de Gruyter.
- Hardy, Donald E. 2004. Collocational Analysis as a Stylistic Discovery Procedure: The Case of Flannery O'Connor's Eyes. *Style* 38(4). 410.
- Hardy, Donald E. 2007. Corpus Stylistics as a Discovery Procedure. In Greg Watson & Sonia Zyngier (eds.), *Literature and stylistics for language learners: theory and practice*, 79–90. Basingstoke: Palgrave Macmillan.
- Hawes, Donald. 2002. *Who's who in Dickens*. 2nd edn. London: Routledge.
- Hawes, Donald. 2007. *Charles Dickens*. London: Continuum.
- Haynes, John. 1995. *Style*. London: Routledge.
- Henrichs, Lotte & Rob Schoonen. 2009. Lexical features of parental academic language input: The effect on vocabulary growth in monolingual Dutch children. In Brian Richards, Michael H. Daller, David D. Malvern, Paul Meara, James Milton & Jeanine Treffers-Daller (eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application*, 1–22. Basingstoke: Palgrave Macmillan.
- Hernández, Purificación S. 2011. The potential of literacy texts in the language classroom: The study of linguistic functions. *Odisea: Revista de Estudios Ingleses*(12). 233–244.
- Hill, David R. 1997. Survey review: Graded readers. *ELT Journal* 51(1). 57–81.
- Hill, Jimmie. 2002. Revising priorities: From grammatical failure to collocational success. In Michael Lewis (ed.), *Teaching collocation: further developments in the lexical approach*, 47–69. Boston, Mass.: Thomson Heinle Language Teaching Publications.
- Hirsh, David & Paul Nation. 1992. What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8(2). 689–96.
- Hockey, Susan. 1994. Evaluating electronic texts in the humanities. *Library Trends* 42(4). 676–693.
- Hockey, Susan. 2000. *Electronic texts in the humanities*. Oxford: Oxford University Press.
- Hoey, Michael. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.
- Hoey, Michael. 2005. *Lexical priming: A new theory of words and language*. London: Routledge.
- Hoover, David L. 1999. *Language and style in the inheritors*. Lanham, Md: University Press of America.
- Hoover, David L. 2002. Frequent word sequences and statistical stylistics. *Literary and Linguistic Computing* 17(2). 157–180.
- Hosman, Lawrence A. 2008. Style and persuasion. In Ulla Fix, Andreas Gardt & Joachim Knape (eds.), *Rhetoric and stylistics: An international handbook*

- of historical and systematic research, vol. 1, 1119–1129. Berlin: Walter de Gruyter.
- Hu, Chunyu. 2015. Using Wmatrix to explore discourse of economic growth. *English Language Teaching* 8(9).
- Huckin, Thomas. 2004. Content analysis: what texts talk about. In Charles Bazerman & Paul A. Prior (eds.), *What writing does and how it does it: An introduction to analyzing texts and textual practices*, 13–32. London: Lawrence Erlbaum Associates, Publishers.
- Hudson, Carla L. & Inge-Marie Eigsti. 2003. The lexical competence hypothesis: A cognitive account of the relationship between vernacularization and grammatical expansion in creolization. *Journal of Pidgin and Creole Languages* 18(1). 1–79.
- Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, Susan. 2006. Corpus Linguistics. In Keith E. Brown & Anne Anderson (eds.), *Encyclopedia of Language & Linguistics*, 234–248. 2nd edn. Amsterdam: Elsevier.
- Hunston, Susan. 2011. *Corpus approaches to evaluation: Phraseology and evaluative language*. London: Routledge.
- Hunt, Daniel & Ronald Carter. 2012. Seeing through the bell jar: Investigating linguistic patterns of psychological disorder. *Journal of Medical Humanities* 33(1). 27–39.
- Hyland, Ken. 2007. English for specific purposes: Some influences and impacts. In Jim Cummins & Chris Davison (eds.), *International handbook of English language teaching*, 391–402. New York: Springer.
- Hyland, Ken. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27(1). 4–21.
- Hymes, Dell H. 1972. On communicative competence. In J. B. Pride & Janet Holmes (eds.), *Sociolinguistics: selected readings*, 269–293. Baltimore: Penguin.
- Jakobson, Roman. 1960. Closing statement: Linguistics and poetics. In Thomas A. Sebeok (ed.), *Style in Language*, 350–377. Cambridge MA: The MIT Press.
- Jaszczolt, Katarzyna. 2005. *Default semantics: foundations of a compositional theory of acts of communication*. Oxford: Oxford University Press.
- Jeffries, Lesley. 2000. Don't throw out the baby with the bathwater: In defence of theoretical eclecticism in Stylistics. *Occasional Papers PALA: The Poetics and Linguistics Association* 12. 3–18.
- Jeffries, Lesley. 2006. *Discovering language: The structure of modern English*. Basingstoke: Palgrave Macmillan.
- Jeffries, Lesley. 2010. *Critical stylistics: The power of English*. Basingstoke: Palgrave Macmillan.

- Jeffries, Lesley & Dan McIntyre. 2010. *Stylistics*. Cambridge: Cambridge University Press.
- Johnson, Keith & Helen Johnson. 1998. *The encyclopedic dictionary of applied linguistics: A handbook for language teaching*. 1st edn. London: Blackwell Publishers.
- Karoly, Adrienn. 2005. The importance of raising collocational awareness in the vocabulary development of intermediate level learners of English. *Eger Journal of English Studies* V. 58–69.
- Kaufmann, Steve. 2003. *The linguist: A personal guide to language learning*. West Vancouver, B.C.: The Linguist Institute.
- Keeseey, Donald (ed.). 1987. *Contexts for criticism*. Mountain View, CA: Mayfield.
- Kelly, Peter. 1991. Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. *International Review of Applied Linguistics in Language Teaching* 29(2). 135–149.
- Kennedy, Graeme. 1998. *An introduction to corpus linguistics*. Reading: Addison Wesley Longman.
- Keshabyan-Ivanova, I. 2014. A computer-assisted analysis of literary texts: a sample study. *Computer modelling & new technologies* 18(6). 59–64.
- Kettemann, Bernhard & Georg Marko. 2004. Can the L in TaLC stand for literature? In Guy Aston, Silvia Bernardini & Dominic Stewart (eds.), *Corpora and language learners*, 169–193. Amsterdam: John Benjamins Publishing Company.
- Knowles, Gerry & Zuraidah M. Don. 2004. The notion of a “lemma”: Headwords, roots and lexical sets. *International Journal of Corpus Linguistics* 9(1). 69–81.
- Kramersch, Claire. 1993. *Context and culture in language teaching*. Oxford: Oxford University Press.
- Kramersch, Claire & Olivier Kramersch. 2000. The avatars of literature in language study. *The Modern Language Journal* 84(4). 553–573.
- Kraus, Jiří. 2008. Structuralist conceptions of style Jiří Kraus. In Ulla Fix, Andreas Gardt & Joachim Knape (eds.), *Rhetoric and stylistics: An international handbook of historical and systematic research*, vol. 1, 1010–1023. Berlin: Walter de Gruyter.
- Laufer, Batia & Paul Nation. 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16(3). 307–322.
- Lazar, Gillian. 1993. *Literature and language teaching: a guide for teachers and trainers*. Cambridge: Cambridge University Press.
- Leaver, Betty L., Madeline E. Ehrman & Boris Shekhtman. 2005. *Achieving success in second language acquisition*. Cambridge: Cambridge University Press.

- Leaver, Betty L. & Boris Shekhtman. 2002. Principles and practices in teaching superior-level language skills: Not just more of the same. In Betty L. Leaver & Boris Shekhtman (eds.), *Developing professional-level language proficiency*, 3–33. Cambridge: Cambridge University Press.
- Leech, Geoffrey. 1992. Corpus linguistics and theories of linguistic performance. In Jan Svartvik (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, 105–122. Berlin: Mouton de Gruyter.
- Leech, Geoffrey N. 1991a. The state of the art in corpus linguistics. In Karin Aijmer & Bengt Altenberg (eds.), *English corpus linguistics: studies in honour of Jan Svartvik*, 105–122. Longman.
- Leech, Geoffrey N. 1991b. Corpora. In Kirsten Malmkjær & James M. Anderson (eds.), *The Linguistics encyclopedia*, 97–106. London: Routledge.
- Leech, Geoffrey N. 1992. Corpora and theories of linguistic performance. In Jan Svartvik (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, 105–122. Berlin: Mouton de Gruyter.
- Leech, Geoffrey N. 2007a. Style in fiction revisited: the beginning of Great Expectations. *Style* 41(2). 117–132.
- Leech, Geoffrey N. 1997b. Introducing corpus annotation. In Roger Garside, Geoffrey N. Leech & Tony McEnery (eds.), *Corpus annotation: Linguistic information from computer text corpora*, 1–18. London: Longman.
- Leech, Geoffrey N. 2008. *Language in literature: style and foregrounding*. London: Routledge.
- Leech, Geoffrey N. & Mick Short. 2007. *Style in fiction: a linguistic introduction to English fictional prose*. 2nd edn. London: Pearson Longman.
- Leech, Geoffrey, Paul Rayson & Andrew Wilson. 2001. *Word frequencies in written and spoken English: based on the British National Corpus*. Harlow: Longman.
- Lewis, Michael. 1993. *The lexical approach: The state of ELT and a way forward*. London: Heinle.
- Lewis, Michael. 2008. *Implementing the lexical approach: Putting theory into practice*. Andover: Heinle, Cengage Learning.
- Litosseliti, Lia. 2010. *Research methods in linguistics*. London: Continuum.
- Long, Michael H. & Steven Ross. 1993. Modifications that preserve language and content. In Makhan L. Tickoo (ed.), *Simplification: Theory and application*, 29–52. Singapore: SEAMEO Regional Language Center.
- Louw, Bill. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and Technology: In honour of John Sinclair*, 155–176. Amsterdam: John Benjamins Publishing.

- Louw, Bill. 1997. The role of corpora in critical literary appreciation. In Anne Wichmann, Steven Fligelstone, Tony McEnery & Gerry Knowles (eds.), *Teaching and language corpora*, 240–51. London: Longman.
- Louw, Bill. 2007. Literary worlds as collocation. In Greg Watson & Sonia Zyngier (eds.), *Literature and stylistics for language learners: theory and practice*, 91–105. Basingstoke: Palgrave Macmillan.
- Louw, Bill & Carmela Chateau. 2010. Semantic prosody for the 21st century: Are prosodies smoothed in academic contexts? A contextual prosodic theoretical perspective. *Proceedings of 10th International Conference on Statistical Analysis of Corpus Study*, 755–764.
- Mackay, Ray. 1996. Mything the point: A critique of objective stylistics. *Language & Communication* 16(1). 81–93.
- Mackay, Ray. 1999. There goes the other foot - a reply to Short et al. *Language and Literature* 8(1). 59–66.
- Mahlberg, Michaela. 2005. *English general nouns: A corpus theoretical approach*. Amsterdam: John Benjamins Publishing Company.
- Mahlberg, Michaela. 2007a. Clusters, key clusters and local textual functions in Dickens. *Corpora* 2(1). 1–31.
- Mahlberg, Michaela. 2007b. Corpus stylistics: bridging the gap between linguistic and literary studies. In Michael Hoey, Michaela Mahlberg & Michael Stubbs (eds.), *Text, Discourse and Corpora: Theory and Analysis*. London: Continuum.
- Mahlberg, Michaela. 2007c. A corpus stylistic perspective on Dickens' Great Expectations. In Marina Lambrou & Peter Stockwell (eds.), *Contemporary Stylistics*, 19–31. London: Continuum.
- Mahlberg, Michaela. 2009. Corpus stylistics and the Pickwickian watering-pot. In Paul Baker (ed.), *Contemporary Corpus Linguistics*, 47–63. London: Continuum.
- Mahlberg, Michaela. 2010. Corpus Linguistics and the Study of Nineteenth-Century Fiction. *Journal of Victorian Culture* 15(2). 292–298.
- Mahlberg, Michaela. 2013. *Corpus stylistics and Dickens' fiction*. Routledge.
- Mahlberg, Michaela. 2014. *Corpus stylistics*. In Michael Burke (ed.), *The Routledge Handbook of Stylistics*, 378–392. London: Routledge.
- Mahlberg, Michaela & Dan McIntyre. 2011. A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction* 4(2). 204–227.
- Mahlberg, Michaela & Catherine Smith. 2010. Corpus Approaches to Prose Fiction: Civility and Body Language in *Pride and Prejudice*. In Dan McIntyre & Beatrix Busse (eds.), *Language and Style*, 449–467. Basingstoke: Palgrave Macmillan.
- Mahlberg, Michaela & Catherine Smith. 2012. Dickens, the suspended quotation and the corpus. *Language and Literature* 21(1). 51–65.
- Maley, Alan & Alan Duff. 1990. *Literature*. Oxford: Oxford University Press.

- Malvern, David & Brian Richards. 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19(1). 85–104.
- Malvern, David, Brian Richards, Ngoni Chipere & Pilar Durán. 2004. *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.
- Matthews, Peter H. 2007. *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.
- Mauranen, Anna. 2004. Spoken corpus for an ordinary learner. In John McHardy Sinclair (ed.), *How to use corpora in language teaching*, 89–105. Amsterdam: John Benjamins Publishing Company.
- McCarthy, Michael & Ronald Carter. 1994. *Language as discourse: Perspectives for language teaching*. Longman Publishing.
- McCrae, John & Urszula Clark. 2005. Stylistics. In Alan Davies & Catherine Elder (eds.), *The Handbook of Applied Linguistics*, 328–346. Oxford: Blackwell Publishing.
- McEnery, Tony & Costas Gabrielatos. 2006. English corpus linguistics. In Bas Arts & April McMahon (eds.), *The handbook of English linguistics*, 33–71. London: Blackwell Publishing.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery, Tony & Andrew Wilson. 2001. *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- McIntyre, Dan. 2008. Integrating multimodal analysis and the stylistics of drama: a multimodal perspective on Ian McKellen's Richard III. *Language and Literature* 17(4). 309–334.
- McIntyre, Dan. 2010. Dialogue and Characterization in Quentin Tarantino's Reservoir Dogs: A Corpus Stylistic Analysis. In Dan McIntyre & Beatrix Busse (eds.), *Language and Style*, 162–182. Basingstoke: Palgrave Macmillan.
- McIntyre, Dan. 2013. The year's work in stylistics 2012. *Language and Literature* 22(4). 333–347.
- McIntyre, Dan. 2014. The year's work in stylistics 2013. *Language and Literature* 23(4). 389–401.
- McIntyre, Dan & Brian Walker. 2010. How can corpora be used to explore the language of poetry and drama? In Anne O'Keeffe & Michael McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*, 516–530. London: Routledge.
- McIntyre, Dan & Brian Walker. 2011. Discourse presentation in Early Modern English writing: A preliminary corpus-based investigation. *International Journal of Corpus Linguistics* 16(1). 101–130.

- McRae, John. 1987. *Words on words: How to write a commentary on a passage of literary prose*. Napoli: Loffredo.
- McRae, John & Roy Boardman. 1984. *Reading between the lines: Integrated language and literature. Teachers' Book*. Cambridge: Cambridge University Press.
- Meara, Paul. 1996. The dimensions of lexical competence. In Gillian Brown, Kirsten Malmkjaer & John Williams (eds.), *Performance and competence in second language acquisition*, 35–52. Cambridge: Cambridge University Press.
- Meara, Paul. 2009. *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins Publishing Company.
- Meara, Paul & Huw Bell. 2001. P-Lex: A simple and effective way of describing the lexical characteristics of short L2 tests. *Prospect* 16(3). 5–19.
- Meara, Paul & James Milton. 2003. *X-Lex: The Swansea Levels Test [CD-ROM]*. Newbury, UK: Express.
- Meyer, Charles F. 2002. *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Miller, J. Hillis. 2002. *On literature*. London: Routledge.
- Miller, Richard E. 2003. Schooling Misery: the ominous threat and the eminent promise of the popular reader. In Tanya Agathocleous & Ann C. Dean (eds.), *Teaching literature: a companion*, 125–138. Basingstoke: Palgrave Macmillan.
- Milojkovic, Marija. 2013a. Is corpus stylistics bent on self-improvement? The role of reference corpora 20 years after the advent of semantic prosody. *Journal of Literary Semantics* 42(1). 59–78.
- Milojkovic, Marija. 2013b. Bill Louw's Contextual Prosodic Theory as the basis of (foreign language) classroom corpus stylistics research. *Research in Corpus Linguistics* 1(1). 47–63.
- Milton, James. 2009. *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Milton, James & Giovanna Donzelli. 2013. The lexicon. In Julia Rogers Herschensohn & Martha Young-Scholten (eds.), *The Cambridge handbook of second language acquisition*, 441–460. Cambridge: Cambridge University Press.
- Montgomery, Martin, Alan Durant, Nigel Fabb, Tom Furniss & Sara Mills. 2007. *Ways of reading: Advanced reading skills for students of English literature*. 3rd edn. London: Routledge.
- Mosteller, Frederick & David L. Wallace. 1964. *Inference and disputed authorship: The Federalist*. Reading, MA: Addison-Wesley.

- Moyer, Alene. 2009. Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In Thorsten Piske & Martha Young-Scholten (eds.), *Input matters in SLA*, 159–174. Bristol: Multilingual Matters.
- Mullany, Louise & Peter Stockwell. 2010. *Introducing English Language: a resource book for students*. London: Routledge.
- Murphy, Sean. 2007. Now I am alone: A corpus stylistic approach to Shakespearian soliloquies. In Costas Gabrielatos, Richard Slessor & J.W. Unger (eds.), *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching*, vol. 1, 67–85. Lancaster: Department of Linguistics and English Language, Lancaster University.
- Naciscione, Anita. 2010. *Stylistic use of phraseological units in discourse*. Amsterdam: John Benjamins Publishing Company.
- Nagy, William. 2007. Metalinguistic awareness and the vocabulary – comprehension connection. In Richard K. Wagner, Andrea E. Muse & Kendra R. Tannenbaum (eds.), *Vocabulary acquisition: Implications for reading comprehension*, 52–77. London: The Guilford Press.
- Nation, I. S. Paul. 1990. *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I. S. Paul. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. Paul. 2004. A study of the most frequent word families in the British National Corpus. In Paul Bogaards & Batia Laufer (eds.), *Vocabulary in a second language: Selection, acquisition, and testing*, 3–14. Amsterdam: John Benjamins Publishing Company.
- Nation, I. S. Paul. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63(1). 59–82.
- Nation, I. S. Paul. 2009. New Roles for L2 Vocabulary? In Wei Li & Vivian Cook (eds.), *Contemporary applied linguistics*, 99–116. London: Continuum.
- Nation, I. S. Paul. 2012. The BNC/COCA word family lists. Available from: http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf (Accessed: 20 October, 2015).
- Nation, I. S. Paul & Stuart Alexander Webb. 2011. *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Nation, Paul. 2001. Using small corpora to investigate learner needs: Two vocabulary research tools. In Mohsen Ghadessy, Alex Henry & Robert L. Roseberry (eds.), *Small corpus studies and ELT: Theory and practice*, 31–45. Amsterdam: John Benjamins Publishing Company.

- Nesselhauf, Nadja. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics* 24(2). 223–242.
- Nesselhauf, Nadja. 2004. *Collocations in a learner corpus*. Amsterdam: John Benjamins Publishing Company.
- O'Halloran, Kieran. 2007a. The subconscious in James Joyce's 'Eveline': a corpus stylistic analysis that chews on the 'Fish hook'. *Language and Literature* 16(3). 227–244.
- O'Halloran, Kieran. 2007b. Corpus-assisted literary evaluation. *Corpora* 2(1). 33–63. doi:10.3366/cor.2007.2.1.33.
- O'Halloran, Kieran. 2012. Performance stylistics: Deleuze and Guattari, poetry and (corpus) linguistics. *International Journal of English Studies* 12(2). 171–199.
- O'Keeffe, Anne. 2006. *Investigating media discourse*. London: Routledge.
- O'Keeffe, Anne, Michael McCarthy & Ronald Carter. 2007. *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Olejniczak, Przemysław. 2006. Phrasal verb idioms and the normative concept of the interlanguage hypothesis. In Janusz Arabski (ed.), *Cross-linguistic influences in the second language lexicon*, 259–272. Clevedon: Multilingual Matters.
- Paquot, Magali. 2010. *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. London: Continuum.
- Picken, Jonathan D. 2007. *Literature, metaphor, and the foreign language learner*. Basingstoke: Palgrave Macmillan.
- Qian, David D. 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning* 52(3). 513–536.
- Radford, Andrew, Martin Atkinson, David Britain, Harald Clahsen & Andrew Spencer. 2009. *Linguistics: An introduction*. 2nd edn. Cambridge: Cambridge University Press.
- Rayson, Paul. 2002. *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. PhD thesis. Lancaster: Lancaster University.
- Rayson, Paul. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics* 13(4). 519–549.
- Reppen, Randi. 2009. English language teaching and corpus linguistics: Lessons from the American National Corpus. In Paul Baker (ed.), *Contemporary Corpus Linguistics*, 204–213. London: Continuum.
- Richards, Jack C. & Richard Schmidt. 2010. *Longman dictionary of language teaching and applied linguistics*. 4th edn. Harlow: Longman.

- Rixon, Shelagh. 2000. Authenticity. In Michael Byram & Adelheid Hu (eds.), *Routledge encyclopedia of language teaching and learning*, 68–69. 1st edn. London: Routledge.
- Romaine, Suzanne. 2010. 19th Century key words, key semantic domains and affect: “In the rich vocabulary of love ‘most dearest’ be a true superlative”. *Studia Neophilologica* 82(1). 12–48.
- Schlicke, Paul (ed.). 2011. *The Oxford companion to Charles Dickens: Anniversary edition*. Oxford: Oxford University Press.
- Schmitt, Diane & Norbert Schmitt. 2005. *Focus on vocabulary: Mastering the academic word list*. 1st edn. White Plains, NY: Pearson Education.
- Schmitt, Norbert. 2000. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Scott, Mike. 1997. PC analysis of key words—and key key words. *System* 25(2). 233–245.
- Scott, Mike. 2011. *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software.
- Scott, Mike & Christopher Tribble. 2006. *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins Publishing Company.
- Seidlhofer, Barbara. 2002. Pedagogy and local learner corpora: Working with learning-driven data. In Sylviane Granger, Joseph Hung & Stephanie Petch-Tyson (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*, 213–234. Amsterdam: John Benjamins Publishing Company.
- Semino, Elena. 2011. Stylistics. In James Simpson (ed.), *The Routledge handbook of applied linguistics*, 541–554. London: Routledge.
- Semino, Elena & Mick Short. 2004. *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.
- Seretan, Violeta. 2011. *Syntax-based collocation extraction*. Dordrecht: Springer.
- Shen, Dan. 2014. Stylistics and narratology. In Michael Burke (ed.), *The Routledge handbook of stylistics*, 191–205. (Routledge Handbooks in English Language Studies). London: Routledge.
- Short, Mick. 1988. *Reading, analysing, and teaching literature*. London: Longman.
- Short, Mick. 1996. *Exploring the language of poems, plays, and prose*. London: Longman.
- Short, Mick, Donald C. Freeman, Willie van Peer & Paul Simpson. 1998. Stylistics, criticism and mythrepresentation again: squaring the circle with Ray Mackay’s subjective solution for all problems. *Language and Literature* 7(1). 39–50.
- Short, Mick & Willie van Peer. 1999. A reply to Mackay. *Language and Literature* 8(3). 269–275.
- Shuy, Roger W. 2001. Discourse analysis in the legal context. In Deborah Schiffrin, Deborah Tannen & Heidi Ehernberg Hamilton (eds.), *The*

- handbook of discourse analysis*, 437–452. (Blackwell Handbooks in Linguistics). London: Blackwell Publishers.
- Sinclair, John McHardy. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Simensen, Aud M. 1987. Adapted readers: How are they adapted? *Reading in a Foreign Language* 4(1). 41–57.
- Simpson, Joellen M. 2000. Topical structure analysis of academic paragraphs in English and Spanish. *Journal of Second Language Writing* 9(3). 293–309.
- Simpson, Paul. 1993. *Language, ideology, and point of view*. London: Routledge.
- Simpson, Paul. 1997. *Language through literature: an introduction*. London: Routledge.
- Simpson, Paul. 2003. *On the discourse of satire: Towards a stylistic model of satirical humor*. Amsterdam: John Benjamins Publishing.
- Simpson, Paul. 2004. *Stylistics: A resource book for students*. London: Routledge.
- Sinclair, John. 2007. Introduction. In Michael Hoey, Michaela Mahlberg & Michael Stubbs (eds.), *Text, discourse and corpora: Theory and analysis*. London: Continuum.
- Sinclair, John McHardy. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, John McHardy. 2004. *Trust the text: Language, corpus and discourse*. London: Routledge.
- Singleton, David. 2000. *Language and the lexicon: An introduction*. London: Arnold.
- Someya, Yasumasa. 1998. English lemma list. Available from: http://www.laurenceanthony.net/resources/wordlists/e_lemma_no_hyphen.zip (Accessed: 10 November, 2014).
- Stelma, Juup. 2010. What is communicative language teaching? In Susan Hunston & David Oakey (eds.), *Introducing applied linguistics: concepts and skills*, 53–59. London: Routledge.
- Stephenson, John. 2009. Learner attitudes towards comprehension-based language learning. In Thorsten Piske & Martha Young-Scholten (eds.), *Input matters in SLA*, 116–136. Bristol: Multilingual Matters.
- Sterkenburg, Piet van (ed.). 2003. *A practical guide to lexicography*. Amsterdam: John Benjamins Publishing Company.
- Stockwell, Peter. 2002. Miltonic texture and the feeling of reading. In Elena Semino & Jonathan Culpeper (eds.), *Cognitive stylistics: language and cognition in text analysis*, 73–94. Amsterdam: John Benjamins Publishing Company.
- Stockwell, Peter. 2006. Language and Literature: Stylistics. In Bas Arts & April McMahon (eds.), *The Handbook of English Linguistics*, 742–758. Oxford: Blackwell Publishing.

- Stockwell, Peter & Sara Whiteley. 2014. Introduction. In Peter Stockwell & Sara Whiteley (eds.), *The Cambridge handbook of stylistics*. Cambridge: Cambridge University Press.
- Stubbs, Michael. 2001. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- Stubbs, Michael. 2005. Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature* 14(1). 5–24.
- Stubbs, Michael. 2010. Three concepts of keywords. In Marina Bondi & Mike Scott (eds.), *Keyness in texts*, 21–42. Amsterdam: John Benjamins Publishing Company.
- Studer, Patrick. 2008. *Historical corpus stylistics: Media, technology and change*. London: Continuum.
- Sucksmith, Harvey Peter. 1970. *The narrative art of Charles Dickens: The rhetoric of sympathy and irony in his novels*. Oxford: Clarendon Press.
- Swaffar, Janet K. 1985. Reading authentic texts in a foreign language: A cognitive model. *The Modern Language Journal* 69(1). 15–34.
- Taylor, Charlotte. 2008. What is corpus linguistics? What the data says. *ICAME Journal*(32). 179–200.
- Thornborrow, Joanna & Shân Wareing. 1998. *Patterns in Language: An Introduction to Language and Literary Style*. London: Routledge.
- Toolan, Michael. 1998. *Language in literature: An introduction to stylistics*. London: Routledge.
- Toolan, Michael. 2006. Top keyword abridgements of short stories: A corpus linguistic resource? *Journal of Literary Semantics* 35(2). 181–194.
- Toolan, Michael. 2008. Narrative progression in the short story: First steps in a corpus stylistic approach. *Narrative* 16(2). 105–120.
- Toolan, Michael. 2011. Teaching the stylistics of prose fiction. In Lesley Jeffries & Dan McIntyre (eds.), *Teaching stylistics*, 178–199. Basingstoke: Palgrave Macmillan.
- Trask, Robert L. 2007. *Language and linguistics: The key concepts*. (Ed.) Peter Stockwell. 2nd edn. London: Routledge.
- VanPatten, Bill & Alessandro G. Benati. 2010. *Key terms in second language acquisition*. London: Continuum.
- Velasco, Daniel G. 2007. Lexical competence and Functional Discourse Grammar. *ALFA: Revista de Lingüística* 51(2). 165–187.
- Walker, Brian David. 2012. *Character and characterisation in Julian Barnes' Talking it over : a corpus stylistic analysis*. Lancaster: Lancaster University Ph.D.
<http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.625457> (1 June, 2016).
- Watkin, Amy S. 2009. *Bloom's how to write about Charles Dickens*. New York: Bloom's Literary Criticism.

- Zaro Vera, Juan J. 1991. Literature as study and resource: The purposes of English literature teaching at university level. *Revista Alicantina de Estudios Ingleses* 4. 163–175.
- Verdonk, Peter. 2002. *Stylistics*. Oxford: Oxford University Press.
- Wagner, Richard K., Andrea E. Muse & Kendra R. Tannenbaum. 2007. Promising avenues for better understanding implications of vocabulary development for reading comprehension. In Richard K. Wagner, Andrea E. Muse & Kendra R. Tannenbaum (eds.), *Vocabulary acquisition: Implications for reading comprehension*, 78–103. London: The Guilford Press.
- Wales, Katie. 2006. Stylistics. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 213–217. 2nd edn. London: Elsevier.
- Wales, Katie. 2011. *A Dictionary of Stylistics*. 3rd edn. London: Pearson Education.
- Waters, Alan. 2009. Advances in materials design. In Michael H. Long & Catherine Doughty (eds.), *The handbook of language teaching*, 311–326. Chichester: Wiley-Blackwell.
- Watson, Greg & Sonia Zyngier (eds.). 2007. *Literature and stylistics for language learners: Theory and practice*. Basingstoke: Palgrave Macmillan.
- Webb, S. & E. Kagimoto. 2011. Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics* 32(3). 259–276.
- Widdowson, Henry G. 1975. *Stylistics and the teaching of literature*. London: Routledge.
- Widdowson, Henry G. 1978. *Teaching language as communication*. Oxford: Oxford University Press.
- Widdowson, Henry G. 1979. *Explorations in applied linguistics*. Oxford: Oxford University Press.
- Widdowson, Henry G. 1992. *Practical stylistics: An approach to poetry*. Oxford: Oxford University Press.
- Widdowson, Henry G. 2008. The novel features of text. Corpus analysis and stylistics. In Andrea Gerbig & Oliver Mason (eds.), *Language, people, numbers corpus linguistics and society*, 293–304. Amsterdam: Rodopi.
- Wilkins, David A. 1972. *Linguistics in language teaching*. London: Arnold.
- Wolf, Shelby A. 2014. *Interpreting Literature With Children*. London: Lawrence Erlbaum Associates, Publishers.
- Wynne, Martin. 2005. Stylistics: corpus approaches. Available from: http://www.pala.ac.uk/uploads/2/5/1/0/25105678/corpora_stylistics.pdf (Accessed: 02 January, 2015).
- Xiao, Richard. 2009. Theory-driven corpus research: Using corpora to inform aspect theory. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 987–1008. Berlin: Walter de Gruyter.

Appendices

Appendix 1.1: Function Words by Paul Nation

Function words usually have high frequency in corpus analysis (see Nation 2001: Appendix 6)

a	beside	even	have
about	between	ever	haven't
above	beyond	every	having
after	billion	everybody	he
again	billionth	everyone	he'd
ago	both	everything	he'll
all	but	everywhere	hence
almost	by	except	her
along	can	far	here
already	can't	few	hers
also	could	fewer	herself
although	couldn't	fifteen	he's
always	did	fifteenth	him
am	didn't	fifth	himself
among	do	fiftieth	his
an	does	fifty	hither
and	doesn't	first	how
another	doing	five	however
any	done	for	hundred
anybody	don't	fortieth	hundredth
anything	down	forty	I
anywhere	during	four	I'd
are	each	fourteen	if
aren't	eight	fourteenth	I'll
around	eighteen	fourth	I'm
as	eighteenth	from	in
at	eighth	get	into
back	eightieth	gets	is
be	eighty	getting	isn't
been	either	got	it
before	eleven	had	its
being	eleventh	hadn't	it's
below	else	has	itself
beneath	enough	hasn't	I've

last	others	that	up
less	ought	that's	us
many	oughtn't	the	very
may	our	their	was
me	ours	theirs	wasn't
might	ourselves	them	we
million	out	themselves	we'd
millionth	over	then	we'll
mine	quite	thence	were
more	rather	there	we're
most	round	therefore	weren't
much	second	these	we've
must	seven	they	what
mustn't	seventeen	they'd	when
my	seventeenth	they'll	whence
myself	seventh	they're	where
near	seventieth	third	whereas
near	seventy	thirteen	which
nearby	shall	thirteenth	while
nearly	shan't	thirtieth	whither
neither	she	thirty	who
never	she'd	this	whom
next	she'll	thither	whose
nine	she's	those	why
nineteen	should	though	will
nineteenth	shouldn't	thousand	with
ninetieth	since	thousandth	within
ninety	six	three	without
ninth	sixteen	thrice	won't
no	sixteenth	through	would
nobody	sixth	thus	wouldn't
none	sixtieth	till	yes
noone	sixty	to	yesterday
nor	so	today	yet
not	some	tomorrow	you
nothing	somebody	too	you'd
now	someone	towards	you'll
nowhere	something	twelfth	your
of	sometimes	twelve	you're
off	somewhere	twentieth	yours
often	soon	twenty	yourself
on	still	twice	yourselves
once	such	two	you've
one	ten	under	
only	tenth	underneath	
or	than	unless	
other	that	until	

Appendix 4.1 Texts Manually Inserted into the DCC

The following texts were inserted manually in DCC; either because they were not included in the electronic copies that I found available in Project Gutenberg and other complementary resource e.g. archive.org or because they were not found in any digital format which then arise the need to key the text in.

The Early Closing Movement

[The Student and Young Man's advocate, January 1845]

TO THE COMMITTEE OF THE
METROPOLITAN DRAPERS' ASSOCIATION.

GENTLEMEN,

I beg to assure you, that it gives me great satisfaction to have the honour of enrolling my name among the vice-presidents of your association.

My engagement will not permit; I regret to say, of my attending your meeting at the Hanover square rooms, on Monday evenings. But, though absent in the body, I am with you in the spirit there and always. I believe that the objects you have in view, are not of greater importance to yourselves than to the welfare and happiness of society – in general; to whom the comfort, happiness, and intelligence of that large class of industrious persons whose claims you advocate, is if rightly understood, a matter of the highest moment and loftiest concern.

I understand the late-hour system to be a means of depriving very many young men of all reasonable opportunities of self-culture and improvement, and of making their labour irksome, weary, and oppressive. I understand the early-hour system to be a means of lightening their labour without disadvantage to any body or any thing, and of enabling them to improve themselves, as all rational creatures are intended to do, and have a right to do; and therefore I hold that there is no more room for choice or doubt between the two, than there is between good and bad, or right and wrong.

I am, Gentlemen,

Your faithful Servant,

CHARLES DICKENS

DEVONSHIRE TERRACE,
York Gate, Regent's Park,
28th march 1844.

Comment:

This letter as in *vol. XXXIV: 388-389, National Edition* was not found in any electronic format and thus was inserted manually. This letter appears in *The Pilgrim Edition of the Letters of Charles Dickens: Volume 4. 1844-1846* on page 88. It also appears in *The Selected Letters of Charles Dickens* by Jenny Hartley entitled as *To The Committee Of The Metropolitan Drapers' Association*. Both *The Pilgrim Edition of the Letters of Charles*

Dickens and *The Selected Letters of Charles Dickens* were published by Oxford University Press.

All the following texts were inserted manually and located the National Edition of Charles Dickens's works.

DRAMATIC RIGHTS IN FICTION
[LETTER TO *The Times*, JANUARY 12, 1861]

SIR,

I shall feel greatly obliged to you if you will allow me to make known to theatrical managers, through your columns, that I believe it is in the power of any English writer of fiction legally to prevent any work of his from being dramatised or adapted for the stage without his consent, and that I have taken measures for the assertion of this right in my own case, and intend to try it with whomsoever may violate it.

It happened but yesterday that I had, in conjunction with Mr. Wilkie Collins, very unwillingly to assert this principle in defence of a joint production of ours against the proprietor of the Britannia Theatre. In a most frank and honest manner he immediately withdrew an announced piece on the night of its intended production, and when his audience were assembled. As I had no earlier opportunity of giving him notice of my intention to uphold the right of authors, and as I inconvenienced a gentleman for whom I have a respect, with great respect, with great reluctance, and should be exceedingly sorry to do the like in any other ease, perhaps you will find space for this letter.

Faithfully yours,
Charles Dickens

GAD'S HILL, January 8.

Vol. XXXIV: 451-452.

THE EARTHQUAKE FELT IN ENGLAND
[LETTER TO *The Times*, JANUARY 12, 1861]

SIR,

As you may think any accurate observation of the shock of earthquake which was felt in various parts of England last Tuesday morning worth publishing, I send you mine.

I was awakened by a violent swaying of my bedstead from side to side, accompanied by a singular heaving motion. It was exactly as if some great beast had been crouching asleep under the bedstead and were now shaking itself and trying to rise. The time by my watch was twenty minutes past three, and I suppose the shock to have lasted nearly a minute. The bedstead, a large iron one, standing nearly north and south, appeared to me to be the only piece of furniture in the room that was heavily shaken. Neither the doors nor the windows rattled, though they rattle enough in windy weather, this house standing alone, on high ground, in the

neighbourhood of two great rivers. There was no noise. The air was very still, and much warmer than it had been in the earlier part of the night. Although the previous afternoon had been wet, the glass had not fallen. I had mentioned my surprise at its standing near the letter 'i' in 'Fair,' and having a tendency to rise. It is recorded in the second volume of the Philosophical Transactions that the glass stood high at Oxford when an earthquake was felt there in September, 1683.

Your faithful servant,
CHARLES DICKENS.

GAD'S HILL, Place, October 7.

Vol. XXXIV: 452-453.

HISTORY OF 'PICKWICK'
[The Athenæum, March 31, 1866]

**Gad's Hill, Place,
March 28, 1866**

As the author of the Pickwick papers (and of one or two other books), I send you a few facts, and no comments, having reference to a letter signed 'R. Seymour,' which in your editorial discretion you publish last week.

Mr. Seymour the artist never originated, suggested, or in any way had to do with, save as illustrator of what I devised, an incident, a character (except the sporting tastes of Mr. Winkle), a name, a phrase, or a word, to be found in the *Pickwick Papers*. I never saw Mr. Seymour's handwriting, I believe, in my life.

I never even saw Mr. Seymour but once in my life, and that was within eight-and-forty hours of his untimely death. Two persons, both still living, were present on that short occasion.

Mr. Seymour died when only the first twenty-four printed pages of the *Pickwick Papers* were published; I think before the next three or four pages were completely written; I am sure before one subsequent line of the book was invented.

In the preface to the Cheap Edition of the *Pickwick Papers*, publish in October 1847, I thus described the origin of that work: 'I was a young man of three-and-twenty, when the present publishers, attracted by some pieces I was at that time writing in the Morning Chronicle newspaper (of which one series had lately been collected and published in two volumes, illustrated by my esteemed friend Mr. George Cruikshank), waited upon me to propose a something that should be published in shilling numbers—then only known to me, or, I believe, to anybody else, by a dim recollection of certain interminable novels in that form, which used, some five-and-twenty years ago, to be carried about the country by pedlars, and over some of which I remember to have shed innumerable tears, before I served my apprenticeship to Life. The idea propounded to me was that the monthly something should be a vehicle for certain plates to be executed by Mr. Seymour, and there was a nation either on the part of that admirable humorous artist, or of my visitor (I forget which), that a 'Nimrod Club,' the members of which were to go out shooting, fishing and so forth, and getting themselves into difficulties through their want of dexterity,

would be the best means of introducing these. I objected, on consideration, that although born and partly bred in the country, I was no great sportsman, except in regard of all kinds of locomotion ; that the idea was not novel, and had been already much used ; that it would be infinitely better for the plates to arise naturally out of the text ; and that I should like to take my own way, with freer range of English scenes and people, and was afraid I should ultimately do so in any case, whatever course I might prescribe to myself at starting. My views being deferred to, I thought of Mr. Pickwick, and wrote the first number, from the proof sheets of which Mr. Seymour made his drawing of the Club, and that happy portrait of its founder, by which he is always recognised, and which may be said to have made him a reality. I connected Mr. Pickwick with a club because of the original suggestion, and I put in Mr. Winkle I expressly for the use of Mr. Seymour. We started with a number of twenty-four pages instead of thirty-two, and four illustrations in lieu of a couple. Mr. Seymour's sudden and lamented death before the second number was published, brought about a quick decision upon a point already in agitation; the number became one of thirty-two pages with two illustrations, and remained so to the end.'

In July, 1849, some incoherent assertion made by the widow of Mr. Seymour, in the course of certain endeavours of hers to raise money, induced me to address a letter to Mr. Edward Chapman, then the only surviving business-partner in the original firm of Chapman and Hall, who first published the Pickwick Papers, requesting him to inform me in writing whether the foregoing statement was correct.

In Mr. Chapman's confirmatory answer, immediately written, he reminded me that I had given Mr. Seymour more credit than was his due. 'As this letter is to be historical,' he wrote, 'I may as well claim what little belongs to me in the matter, and that is, the figure of Pickwick. Seymour's first sketch,' make from the proof of my first chapter, 'was of a long, thin man. The present immortal one he made from my description of a friend of mine at Richmond.'

CHARLES DICKENS.

[The Athenaeum, APRIL 7, 1866]

There is a verbal mistake in the letter I addressed to you last week, as it is printed in your journal. In the fifth paragraph, the passage, 'I think, before the next three or four pages were completely written,' should stand, 'I think before the next twenty-four pages were completely written.'

CHARLES DICKENS.

Vol. XXXIV: 469-470.

No. 16 Wellington Street North, Strand,
Monday, Twelfth May, 1851.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – My only hesitation on the matter is this: I apprehend that the Duke in his great generosity intends to give a sort of supper to the whole party. I infer this from his so particularly desiring to know their number. *Now, I have already given him the list;* and he is so delicate that he would not even ask Landseer without

first asking me. Under these circumstances, I feel the introduction of a stranger like Mr. Ward's brother—Mr. Ward and his wife being already on the list—a kind of difficulty; but I do not like to refuse compliance with any wish of my faithful and attached valet, whom I greatly esteem. I therefore merely mention this and send him the order.

I have been here all day, and am covered with Sawdust.

Faithfully yours always,

Vol. XXXVII: 267.

TAVISTOCK HOUSE, Twenty-third December, 1852.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, — I am suddenly laid by the heels in consequence of Wills having gone blind without any notice—I hope and believe from mere temporary inflammation. This obliges me to be at the office all day to-day, and to resume my attendance there to-morrow. But if you will come there to-morrow afternoon—say at about three o'clock—I think we may forage pleasantly for a dinner in the City, and then go and look at Christmas Eve in Whitechapel, which is always a curious thing.

Vol. XXXVII: 306.

TAVISTOCK HOUSE, Tuesday, January Eighteenth, 1853.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, — If you should be disposed to revel in the glories of the eccentric British Drayma, on Saturday evening, I am the man to join in so great a movement. My money is to be heard of at the Bar of the Household Words at five o'clock on that afternoon.

Gin Punch is also to be heard of at the Family Arms, Tavistock, on Sunday next at five, when the National Sparkler will be prepared to give Lithers a bellyful if he means anything but Bounce.

I have been thinking of the Italian project, and reducing the time to two months — from the 20th October to the 20th December — see the way to a trip that shall really not exclude any foremost place, and be reasonable too. Details when we meet.

Ever faithfully.

Vol. XXXVII: 310.

BOULOGNE, Thirtieth June, 1853. Thursday.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, — I am very sorry indeed to hear so bad an account of your illness, and had no idea it had been so severe. I can't help writing (though most unnecessarily I hope) to say that you can't get well too soon; and that I warrant the

pure air, regular hours, and perfect repose of this place to bring you round triumphantly. You have only, when you are sufficiently restored, to defy the Doctor and all his works, to write me a line naming your day and hour. My friend Lord Wilmot will then be found at the Custom House.

Ward's account of me was the true one. I was thoroughly disabled — in a week — and doubt if you would have known me. But I recovered with surprising quickness, positively insisting on coming here, against all advice but [Dr.] Elliotson's — and got to work next day but one as if nothing had happened.

And what was the matter with me? Sir — I find this reads like Dr. Johnson directly — Sir, it was an old, afflicted

KIDNEY,

once the torment of my childhood, in which I took cold.

Vol. XXXVII: 317.

TAVISTOCK HOUSE, Monday, Twenty-fourth April, 1854.

MY DEAR COLLINS, — I met the Colonel at the Water Colours on Saturday, and asked him if he would assist in scattering the family dinner next Sunday at half past five, as usual. Will you join us, Sir?

Beaucourt's house above the Moulineaux, on the top of the hill—free and windy—not so bijou-ish, but larger rooms, and possessing a back gate and a field, secured by the undersigned contracting party from the middle of June to the middle of October. I hope you will write the third volume of 'that' book there.

[Chauncey Hare] Townshend coming to town on the 12th of May. Pray Heaven he may not have another choral birthday, and another frolicultural 1 cauliflower.

Ever faithfully, C. D.

1 I think the word is a bold one. It is intended for floricultural.—C. D. [footnote].

Vol. XXXVII: 363.

TAVISTOCK HOUSE, Sixth June, 1854.

Mr. W. Wilkie Collins.

MY DEAR COLLINS:

Form of trip appointment, in compliance with Act of Parliament. Victoria, cap. 7, sec. 304	Day,	Thursday.
	Hour,	Quarter past 11 A.M.
	Place,	Dover Terminus, London Bridge

	Destination, Tunbridge Wells.
	Description of Railway Qualification, Return Ticket
(Signed) Entd.	Charles Dickens.

Vol. XXXVII: 365.

TAVISTOCK HOUSE, Seventh June, 1854.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – Mark has got something in his foot—which is not Gout, of course, though it has a family likeness to that disorder—which he thinks will disable him to-morrow. Under these circumstances, and as this inclement season of summer has set in with so much severity, I think it may be best to postpone our expedition. Will you take a stroll on Hampstead Heath, and dine here on Sunday instead? And if yes, will you be here at two?

Ever faithfully.

Vol. XXXVII: 365-366.

TAVISTOCK HOUSE, Sunday, Seventeenth December, 1854.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – Many thanks for your note. As I rode home in the hansom, that Gravesend night, one or two doubts arose in my mind respecting the Cowell facts and before breakfast on the following morning I wrote to Mark, begging him to say nothing to Jerrold from me until I should have satisfied my mind. I am so sorry at heart for the working-people when they get into trouble, and have their wretched arena chalked out for them with such extraordinary complacency by small political economists, that I have a natural impulse upon me, almost always, to come to the rescue—even of people I detest, if I believe them to have been true to these poor men. I am away to Reading to read the *Carol*, and to Sherborne, and, after Christmas Day, to Bradford, in Yorkshire. The thirtieth will conclude my public appearances for the present season, and then I hope we shall have some Christmas diversions here. I have got the children's play into shape, so far as the Text goes (it is an adaptation of *Fortunio*), but it has not been 'on the stage' yet. Mark is going to do the Dragon—with a practicable head and tail.

Ever
yours.

Vol. XXXVII: 380-381.

TAVISTOCK HOUSE, Christmas Eve, 1854.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – Here is a Part in *Fortunio*—dozen words—but great Pantomime opportunities—which requires a first-rate old stager to devour Property Loaves. Will you join the joke and do it? Gobbler, one of the seven gifted servants, is the Being ‘to let.’ There is an eligible opportunity of making up dreadfully greedy.

I am going to read the piece to the children next Tuesday, at half-past two. We shall rehearse it at the same hour every day in the following week—dress rehearsal on Saturday night, the 6th; night of performance, Monday, the 8th.

I am just come back from Reading and Sherborne, and go to Bradford on Wednesday morning, returning next day.

If you should chance to be disengaged to-day, here we are—Pork, with sage and onions, at half past five.

Ever faithfully.

Vol. XXXVII: 381.

TAVISTOCK HOUSE, Sunday, Fourth March, 1855.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, –

I have to report another failure on the part of our friend ‘Williams’ last night. He so confounded an enlightened British audience at the Standard Theatre on the subject of *Antony and Cleopatra* that I clearly saw them wondering, towards the end of the Fourth Act, when the play was going to begin.

A man much heavier than Mark (in the actual scale, I mean), and about twenty years older, played Cæsar. When he came on with a map of London—pretending it was a scroll and making believe to read it—and said, ‘He calls me Boy’—a howl of derision arose from the audience which you probably heard in the Dark, without knowing what occasioned it. All the smaller characters, having their speeches much upon their minds, came in and let them off without the slightest reference to cues. And Miss Glyn, in some entirely new conception of her art, ‘read’ her part like a Patter song—several lines on end with the rapidity of Charles Mathews, and then one very long word. It was very brightly and creditably got up, but (as I have said) ‘Williams’ did not carry the audience, and I don’t think the Sixty Pounds a week will be got back by the Manager.

You will have the goodness to picture me to yourself—alone—in profound solitude—in an abyss of despair—ensconced in a small Managerial Private Box in the very centre of the House—frightfully sleepy (I had a dirty steak in the City first, and I think they must have put Laudanum into the Harvey’s sauce), and played at, point-blank, by the entire strength of the company. The horrors in which I constantly woke up, and found myself detected, you will imagine. The gentle Glyn, on being called for, heaved her snowy bosom straight at me, and the box-keeper informed me that the Manager who brought her on would ‘have the honour of stepping round directly.’ I sneaked away in the most craven and dastardly manner, and made an utterly false representation that I was coming back again.

If you will give me one glass of hot gin-and-water on Thursday or Friday evening, I will come up about eight o'clock with a cigar in my pocket and inspect the Hospital. I am afraid this relaxing weather will tell a little faintly on your medicine, but I hope you will soon begin to see land beyond the Hunterian Ocean.

I have been writing and planning and making notes over an immense number of little bits of paper—and I never can write legibly under such circumstances.

Always cordially yours.

Vol. XXXVII: 396-397.

TAVISTOCK HOUSE, Monday, Nineteenth March, 1855.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – I have read the two first portions of *Sister Rose* with the very greatest pleasure. An excellent story, charmingly written, and shewing everywhere an amount of pains and study in respect of the art of doing such things that I see mighty seldom.

If I be right in supposing that the brother and sister are concealing the husband's mother, then will you look at the closing scene of the second part again, and consider whether you cannot make the indication of that circumstance a little more obscure—or, at all events, a little less emphatic; as by Rose's only asking her brother once for leave to tell her husband, or some slight alteration of that kind? The best way I know of strengthening the interest and hitting this point would be the introduction or mention, in the first instance, *of some one other person who might* (in the reader's divided thoughts) be the concealed person, and of whom the husband might have a latent dislike or jealousy—as a friend of the brother's. But this might involve too great a change.

If, on the other hand, it be not the mother who is visited, then it is clear that you have altogether succeeded as it stands, and have entirely misled me.

How are you getting on? Shall you be up to a day at Ashford to-morrow week? I shall be able to frank you down and up the Railway on the solemn occasion. Mark (whose face is at present enormous) is going, and Wills will tell us the story of the Bo'sen, whose artful chaff, in that sparkling dialogue, played the Devil with T. Cooke.

Talking of which feat, I wish you could have seen your servant last Wednesday beleaguer the Literary Fund. They got so bothered and bewildered that I expected to see them all fade away under the table; and the outsiders laughed so irreverently whenever I poked up the chairman that it was quite a facetious business. Virtually, I consider the thing done. You may believe that I am not about to let go, and the effect has far and far exceeded my expectations already. Mark is full of the subject and will tell you all about it. . . .

What is Mr. Pigott's address? I want to leave a card for him.

Ever faithfully.

Vol. XXXVII: 397-398.

TAVISTOCK HOUSE, Saturday, Twenty-fourth March, 1855.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – I am charmed to hear of the great improvement, and really hope now that you are beginning to see land.

The train (an express one) leaves London Bridge Station on Tuesday at half past 11 in the forenoon. Fire and comfort are ordered to be in readiness at the Inn at Ashford. We shall have to return at half past 2 in the morning—getting to town before 5—but the interval between the Reading and the Mails will be spent by what would be called in a popular musical entertainment ‘the flick o’ our ain firesides’—which reminds me to observe that I am dead sick of the Scottish tongue in all its moods and tenses.

You have guessed right! The best of it was that she [Mrs. Gaskell] wrote to Wills, saying she must particularly stipulate not to have her proofs touched, ‘even by Mr. Dickens.’ That immortal creature had gone over the proofs [*North and South*] with great pains—had of course taken out the stiflings—hard-plungings, lungeings, and other convulsions—and had also taken out her weakenings and damagings of her own effects. ‘Very well,’ said the gifted Man, ‘she shall have her own way. But after it’s published shew her this Proof, and ask her to consider whether her story would have been the better or the worse for it.’

When you see Millais, tell him that if he would like a quotation for his fireman picture there is a very suitable and appropriate one to be got from Gay’s *Trivia*. . . .

Ever
yours,

I dined with an old General yesterday, who went perfectly mad at dinner about the *Times*—exudations taking place from his mouth while he denied all its statements, that were partly foam, and partly turbot with white sauce. He persisted, likewise, in speaking of that Journal as ‘Him.’

Vol. XXXVII: 398-399.

TAVISTOCK HOUSE, Wednesday, Fourth April, 1855.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – I have read the article in the *Leader* on Napoleon’s reception in England with great pleasure and entire concurrence. I think it is forcible and just, and yet states the real case with great moderation. Not knowing of it, I had been speaking to its author on that very subject in the Pit of the Olympic on Saturday night.

And, by-the-bye, as the Devil would have it (for I assume that he is always up to something, and that everything is his fault—I being, as you know, evangelical), I mislaid your letter with Mr. Pigott’s address in it, and ‘didn’t like’ to ask him for it. Do, like an amiable, corroded hermit, send me that piece of information again.

I hope the medical authorities will not—as I may say—cut your nose off to be revenged on your face. You might want it at some future time. It is but natural that the Doctor should be irritated by so much opposition—still, isn’t the offending feature in some sort a man and a brother?

The Pantomime was amazingly good, and it really was a comfortable thing to see all conventional dignity so outrageously set at naught. It was astonishingly well done, and extremely funny. Not a man in it who wasn't quite as good as the Humbugs who pass their lives in doing nothing else. I observed at the Fund Dinner that the actors are in the same condition about it as they were when we played. Idiots!

May the Spring advance with rosy foot, and the voice of the Turtle be shortly heard in the land.

Ever faithfully.

Vol. XXXVII: 403.

TAVISTOCK HOUSE, Sunday, Fifteenth April, 1855.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – Hurrah!

I shall be charmed to see you once more in a Normal state, and propose Friday next for our meeting at the Garrick, at a quarter before 5. We will then proceed to the Ship and Turtle.

I fell foul of Wills yesterday, for that in 'dealing with' the second part of your story [*Sister Rose*] he had not (in two places) 'indoctrinated' the Printer with the change of name. He explained to me that on the whole, and calmly regarding all the facts from a politico-economical point of view, it was a more triumphant thing to have two mistakes than none—and, indeed, that, philosophically considered, this was rather the object and province of a periodical.

Vol. XXXVII: 405.

TAVISTOCK HOUSE, Thursday, Twenty-first May, 1855.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – Lemon assures me that the Parts and Prompt book are to arrive to-day. Why they have not been here two days I cannot for the life of me make out. In case they *do* come, there is a good deal in the way of clearing the ground that you and I may do before the first Rehearsal. Therefore, will you come and dine at 6 to-morrow (Friday) and give the evening to it?

Faithfully ever.

Vol. XXXVII: 409.

TAVISTOCK HOUSE, Saturday Morning, June Ninth, 1855.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – I have had a communication from Stanfield since we parted last night, to the effect that he must have the Stage entirely to himself and his men on Thursday Night. I therefore write round to all the company, to remind them that Monday is virtually our last Rehearsal, and that we shall probably have to do your Play twice on that precious occasion.

Ever heartily
yours.

Vol. XXXVII: 409.

49 CHAMPS ELYSÉES, Thirtieth January, 1856, Wednesday.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, –

I hope you are ‘out of the wood, and holloaing.’

I purpose coming to town either on Monday or Tuesday night, and returning (if convenient to you), on the following Sunday or Monday. I will write to you as soon as I arrive, and arrange for our devoting an early evening (I should like Wednesday next) to letting our united observation with extended view ‘survey mankind from China to Peru.’ On second thoughts, shall we appoint Wednesday now? Unless I hear from you to the contrary, I will expect you at *Household Words* at five that day. Ever faithfully (working hard).

Vol. XXXVII: 434.

49 CHAMPS ELYSÉES, PARIS. Sunday, Twenty-fourth February, 1856.

Mr. W. Wilkie Collins.

MY DEAR COLLINS, – The Post still coming in to-day without any intelligence from you, I am getting quite uneasy. From day to day I have hoped to hear of your recovery, and have forborne to write, lest I should unintentionally make the time seem longer to you. But I am now so very anxious to know how you are that I cannot hold my hand any longer. So pray let me know by return. And if you should unhappily be too unwell to write yourself, pray get your brother to represent you. I cannot tell you how unfortunate I feel this to be, or how disconsolately I look at the uninhabited Pavilion.

Ever faithfully.

Vol. XXXVII: 436.

XII

London, February 6, 1850.

PUBLIC HEALTH MEETING

[At a public meeting held at Freemasons' Hall under the auspices of the Metropolitan Sanitary Association to consider the question of the Public Health of the Metropolis, the Bishop of London presiding, Mr. Dickens in seconding the third resolution of the meeting regarding the necessity of efficient sanitary precautions against disease made the following speech:]

That the object was to bring the metropolis within the provisions of the Public Health Act, most absurdly and monstrosly excluded from its operation. The object was to diminish an amount of suffering and waste of life which would be a disgrace to a heathen land, to atone for long years of neglect, of which they bad all, to a greater or less extent, been guilty, and to redress a most grievous and cruel injustice. It was a common figure of speech, whenever anything important was left out of any great

scheme, to say it was the tragedy of "Hamlet" with the part of Hamlet left out; but the existence of a Public Health Act, with the metropolis excluded from its operation, suggested to him something even more sad, and that was a representation of the tragedy of "Hamlet" with nothing in it but the gravedigger. This was a state of things which must not last. Every year more than 13,000 unfortunate persons died unnaturally and prematurely around us.¹

Infancy was made stunted, ugly, and full of pain — maturity made old — and old age imbecile; and pauperism made hopeless every day. They claimed for the metropolis of a Christian country that this should be remedied, and that the capital should set an example of humanity and justice to the whole empire. Of the sanitary condition of London at present, he believed it would be almost impossible to speak too ill. He knew of many places in it unsurpassed in the accumulated horrors of their long neglect by the dirtiest old spots in the dirtiest old towns, under the worst old governments in Europe. Among persons living in such a state of civilised society as that in which they lived, there must be contrasts of rank and intelligence, and greater contrasts in reference to wealth and comfort; but he believed that no greater contrasts between wealth and poverty existed in any part of the world than in this metropolis. The principal objectors to the improvements proposed were divided into two classes. The first consisted of the owners of small tenements, men who pushed themselves to the front of boards of guardians and parish vestries, and were clamorous about the rating of their property; the 'other class was composed of gentlemen more independent and less selfish, who had a very weak leaning to the words self-government. The first class generally proceeded upon the supposition that the compulsory improvement of their property, when exceedingly defective, would be very expensive. This was a great mistake, for nothing was cheaper than good sanitary improvement, as had been shown in the case of "Jacob's Island," which he had described in a work of fiction, and where the improvements could be made at a cost of less than the price of a pint of porter, or two glasses of gin a week, to each inhabitant. With regard to the principle of self-government and that what was done in the next parish was no business of theirs, he should begin to think there was something in it when he found any court or street able to keep its diseases within its own bounds, or any parish able to make out the bounds of its own diseases, keeping exclusively to itself its own fever, smallpox, consumption, and pestilence, just as it maintained its own beadles and its fire-engines. Until that time arrived, and so long as he breathed the same air, lived upon the same soil, and under the same sun, he should consider the health and sickness of that parish as being most decidedly his business, and he would endeavour to force it to be cleanly, and would place it under the control of a general board for the general good. The right reverend chairman had referred to the charge made by thoughtless and inconsiderate people, that the poor liked to be dirty and to lead degraded lives. If this charge were true, it would only present another proof that we are living in a most unnatural state of society: but it is no more true than it was true that when they first had baths they would not bathe,

¹ Taking the last four year, more than 15,000 person have on an average, died prematurely in the metropolis, i.e. the mortality has exceeded 2 per cent. Of the population (corrected for increase) that that amount.

and when they first had washhouses they would not wash. We could not expect to "gather grapes, from thorns or figs from thistles." We could not be surprised if the poor did not very highly estimate the decencies of life when they had no opportunity of being made acquainted with them. The main wonder in connection with the poor was, that they did so soon esteem that which was really for their good, when they had any fair experience of it. No one who had any knowledge of the poor could fail to be deeply affected by their patience and their sympathy with one another-by the beautiful alacrity with which they helped each other in toil, in the day of suffering, and in the hour of death. It hardly ever happened that any case of extreme protracted destitution found its way into the public prints, without our reading at the same time of some ragged Samaritan sharing his last loaf, or spending his last penny to relieve the poor miserable in the room upstairs, or in the cellar underground. It was to develop in these people the virtue which nothing could eradicate, to raise them in the social scale as they should be raised, to lift them from a condition into which they did not allow their beasts to sink, and to cleanse the foul air for the passage of Christianity and education throughout the land that the meeting was assembled. He could not lay it to his heart, nor could he flatter any of those present with the idea, that they were met to praise themselves, for they could claim little merit for each other in such a cause. The object of their assembling, as he regarded it, was simply to help to set that right which was very wrong before God and before man.

Vol. XXXVII: 481.

Appendix 4.2: The Comprehensive List of Charles Dickens's Works

The Works of Charles Dickens. National Edition. Chapman and Hall 1906.

1. **Sketches by Boz I – Illustrative of every-day life and every-day people – Sketches of Yong Gentlemen – Sketches of Young Couples – The Mudfog Papers**
 1. Our Parish
 - 1.CHAP. I. The Beadle. The Parish Engine. The Schoolmaster
 - 2.CHAP. II. The Curate. The Old Lady. The Half-Pay Captain
 - 3.CHAP. III. The Four Sisters
 - 4.CHAP. IV. The Election for Beadle
 - 5.CHAP. V. The Broker's Man Mr. Bung's Narrative
 - 6.CHAP. VI. The Ladies' Societies
 - 7.CHAP. VII. Our Next-Door Neighbour
 2. Scenes
 - 1.CHAP. I. The Streets—Morning
 - 2.CHAP. II. The Streets—Night
 - 3.CHAP. III. Shops and Their Tenants
 - 4.CHAP. IV. Scotland-Yard
 - 5.CHAP. V. Seven Dials
 - 6.CHAP. VI. Meditations in Monmouth Street
 - 7.CHAP. VII. Hackney Coach Stands
 - 8.CHAP. VIII. Doctors' Commons
 - 9.CHAP. IX. London Recreations
 10. CHAP. X. The River
 11. CHAP. XI. Astley's
 12. CHAP. XII. Greenwich Fair
 13. CHAP. XIII. Private Theatres
 14. CHAP. XIV. Vauxhall Gardens by Day
 15. CHAP. xv. Early Coaches
 16. CHAP. XVI. Omnibuses
 17. CHAP. XVII. The Last Cab-Driver, and the First Omnibus Cad
 18. CHAP. XVIII. A Parliamentary Sketch
 19. CHAP. XIX . Public Dinners
 20. CHAP. XX. The First of May
 21. CHAP. XXI . Brokers' and Marine-Store Shops
 22. CHAP. XXII . Gin-Shops
 23. CHAP. XXIII . The Pawnbroker's Shop
 24. CHAP. XXIV . Criminal Courts
 25. CHAP. XXV . A Visit to Newgate
 3. Characters
 - 1.CHAP. I. Thoughts about People

- 2.CHAP. II. A Christmas Dinner
- 3.CHAP. III. The New Year
- 4.CHAP. IV. Miss Evans and the Eagle
- 5.CHAP. V. The Parlour Orator
- 6.CHAP. VI. The Hospital Patient
- 7.CHAP. VII. The Misplaced Attachment of Mr. John Dounce
- 8.CHAP. VIII . The Mistaken Milliner. A Tale of Ambition
- 9.CHAP. IX. THE Dancing Academy
10. CHAP. X. Shabby-Genteel People
11. CHAP. XI. Making a Night of it
12. CHAP. XII. The Prisoners' Van

4. Tales

- 1.CHAP. I. the Boarding-House
 1. Chapter the First
 2. Chapter the Second
- 2.CHAP. II. Mr. Minns and His Cousin
- 3.CHAP. III . Sentiment

2. Sketches by Boz II – Illustrative of every-day life and every-day people – Sketches of Yong Gentlemen – Sketches of Young Couples – The Mudfog Papers

- 4.CHAP. IV. The Tuggses at Ramsgate
- 5.CHAP. V. Horatio Sparkins
- 6.CHAP. VI. The Black Veil
- 7.CHAP. VII. The Steam Excursion
- 8.CHAP. VIII. The Great Winglebury Duel
- 9.CHAP. IX. Mrs. Joseph Porter
10. CHAP. X. A Passage in the Life of Mr. Watkins Tottle
11. CHAP. XI. The Bloomsbury Christening
12. CHAP. XII. the Drunkard's Death
13. SKETCHES OF YOUNG GENTLEMEN
14. The Bashful Young Gentleman
15. The Out-And-Out Young Gentleman
16. The Very Friendly Young Gentleman
17. The Military Young Gentleman
18. The Political Young Gentleman
19. The Domestic Young Gentleman
20. The Censorious Young Gentleman
21. The Funny Young Gentleman
22. The Theatrical Young Gentleman
23. The Poetical Young Gentleman
24. The 'Throwing-off' Young Gentleman
25. The Young Ladies' Young Gentleman
26. Conclusion

SKETCHES OF YOUNG COUPLES

- 1.An urgent remonstrance
- 2.The Young Couple

3. The Formal Couple
4. The Loving Couple
5. The Contradictory Couple
6. The Couple who Dote upon their Children
7. The Cool Couple
8. The Plausible Couple
9. The Nice Little Couple
10. The Egotistical Couple
11. The Couple who Coddle Themselves
12. The Old Couple
13. Conclusion

THE MUDFOG AND OTHER SKETCHES

1. Public Life of Mr. Tulrumbly
2. Full Report of the First Meeting of the Mudfog Association
3. Full Report of the Second Meeting of the Mudfog Association
4. SKETCHES
5. The Pantomime of Life
6. Some Particulars Concerning a Lion
7. Mr. Robert Bolton
8. Familiar Epistle from a Parent to a Child

3. The Pickwick Papers VOL. I

1. ADDRESSES AND PREFACES
2. ADDRESS WHICH APPEARED IN PART X., JANUARY, 1837
3. ADDRESS WHICH APPEARED IN PART XV., JULY, 1837
4. PREFACE TO THE FIRST EDITION
5. PREFACE TO THE FIRST CHEAP EDITION
6. PREFACE TO THE 'CHARLES DICKENS' EDITION
7. The Pickwickians
8. The first Day's Journey, and the first Evening's Adventures; with their Consequences
9. A new Acquaintance. The Stroller's Tale. A disagreeable Interruption, and an unpleasant Encounter
10. A Field-Day and Bivouac. More new Friends. An Invitation to the Country
11. A short one. Showing, among other Matters, how Mr. Pickwick undertook to drive, and Mr. Winkle to ride, and how they both did it
12. An old-fashioned Card-party. The Clergyman's Verses. The Story of the Convict's Return
13. How Mr. Winkle, instead of shooting at the Pigeon and killing the Crow, shot at the Crow and wounded the Pigeon; how the Dingley Dell Cricket Club played All-Muggleton, and how All-Muggleton dined at the Dingley Dell expense: with other interesting and instructive Matters
14. Strongly illustrative of the Position, that the Course of True Love is not a Railway

15. A Discovery and a Chase
16. Clearing up all Doubts (if any existed) of the Disinterestedness of Mr. Jingle's Character
17. Involving another Journey, and an Antiquarian Discovery; Recording Mr. Pickwick's Determination to be present at an Election; and containing a Manuscript of the old Clergyman's
18. Descriptive of a very important Proceeding on the Part of Mr. Pickwick; no less an Epoch in his Life, than in this History
19. Some Account of Eatanswill; of the State of Parties therein; and of the Election of a Member to serve in Parliament for that ancient, loyal, and patriotic Borough
20. Comprising a brief Description of the Company at the Peacock assembled; and a Tale told by a Bagman
21. In which is given a faithful Portraiture of two distinguished Persons; and an accurate Description of a public Breakfast in their House and Grounds: which Public Breakfast leads to the Recognition of an old Acquaintance, and the commencement of another Chapter
22. Too full of Adventure to be briefly described
23. Showing that an Attack of Rheumatism, in some Cases, acts as a Quickener to inventive Genius
24. Briefly illustrative of two Points; –first, the Power of Hysterics, and, secondly, the Force of Circumstances
25. A pleasant Day, with an unpleasant Termination
26. Showing how Dodson and Fogg were Men of Business, and their Clerks Men of pleasure; and how an affecting Interview took place between Mr. Weller and his long-lost Parent; showing also what choice Spirits assembled at the Magpie and Stump, and what a capital Chapter the next one will be
27. In which the Old Man launches forth into his favourite Theme, and relates a Story about a queer Client
28. Mr. Pickwick journeys to Ipswich; and meets with a romantic Adventure with a middle-aged Lady in Yellow Curl Papers
29. In which Mr. Samuel Weller begins to devote his Energies to the Return Match between himself and Mr. Trotter
30. Wherein Mr. Peter Magnus grows jealous, and the middle-aged Lady apprehensive, which brings the Pickwickians within the grasp of the Law
31. Showing, among a variety of pleasant Matters, how majestic and impartial Mr. Nupkins was; and how Mr. Weller returned Mr. Job Trotter's Shuttlecock as heavily as it came. With another Matter, which will be found in its Place
32. Which contains a brief Account of the Progress of the Action of Bardell against Pickwick
33. Samuel Weller makes a Pilgrimage to Dorking, and beholds his Mother-in-law

34. A good-humoured Christmas Chapter, containing an Account of a Wedding, and some other Sports beside: which although in their Way, even as good Customs as Marriage itself, are not quite so religiously kept up, in these degenerate Times
35. The Story of the Goblins who stole a Sexton

4. The Pickwick Papers VOL. II

36. How the Pickwickians made and cultivated the Acquaintance of a couple of nice Young Men belonging to one of the Liberal Professions; how they disported themselves on the Ice; and how their Visit came to a Conclusion
37. Which is all about the Law, and sundry Great Authorities learned therein
38. Describes, far more fully than the Court Newsman ever did, a Bachelor's Party, given by Mr. Bob Sawyer at his Lodgings in the Borough
39. Mr. Weller the elder delivers some Critical Sentiments respecting Literary Composition; and, assisted by his son Samuel, pays a small Instalment of Retaliation to the Account of the Reverend Gentleman with the Red Nose
40. Wholly devoted to a full and faithful Report of the memorable Trial of Bardell against Pickwick
41. In which Mr. Pickwick thinks he had better go to Bath; and goes accordingly
42. The chief Features of which, will be found to be an authentic Version of the Legend of Prince Bladud, and a most extraordinary Calamity that befell Mr. Winkle
43. Honourably accounts for Mr. Weller's Absence, by describing a Soiree to which he was invited and went; also relates how he was intrusted by Mr. Pickwick with a Private Mission of Delicacy and Importance
44. How Mr. Winkle, when he stepped out of the Frying-pan, walked gently and comfortably into the Fire
45. Mr. Samuel Weller, being intrusted with a Mission of Love, proceeds to execute it; with what Success will hereinafter appear
46. Introduces Mr. Pickwick to a new and not uninteresting Scene in the great Drama of Life
47. What befell Mr. Pickwick when he got into the Fleet; what Prisoners he saw there; and how he passed the Night
48. Illustrative, like the preceding one, of the old Proverb, that Adversity brings a Man acquainted with strange Bedfellows. Likewise containing Mr. Pickwick's extraordinary and startling announcement to Mr. Samuel Weller
49. Showing how Mr. Samuel Weller got into Difficulties
50. Treats of divers little Matters which occurred in the Fleet, and of Mr. Winkle's mysterious Behaviour; and shows how the poor Chancery Prisoner obtained his Release at last

51. Descriptive of an affecting Interview between Mr. Samuel Weller and a Family Party. Mr. Pickwick makes a Tour of the diminutive World he inhabits, and resolves to mix with it, in Future, as little as possible
52. Records a touching Act of delicate Feeling, not unmixed with Pleasantry, achieved and performed by Messrs. Dodson and Fogg
53. Is chiefly devoted to Matters of Business, and the temporal Advantage of Dodson and Fogg. Mr. Winkle reappears under extraordinary circumstances. Mr. Pickwick's Benevolence proves stronger than his Obstinacy
54. Relates how Mr. Pickwick, with the Assistance of Samuel Weller, essayed to soften the Heart of Mr. Benjamin Allen, and to mollify the Wrath of Mr. Robert Sawyer
55. Containing the Story of the Bagman's Uncle
56. How Mr. Pickwick sped upon his Mission, and how he was reinforced in the Outset by a most unexpected Auxiliary
57. In which Mr. Pickwick encounters an old Acquaintance. To which fortunate Circumstance the Reader is mainly indebted for Matter of thrilling interest herein set down, concerning two great Public Men of might and power
58. Involving a serious Change in the Weller Family and the untimely Downfall of the red-nosed Mr. Stiggins
59. Comprising the final Exit of Mr. Jingle and Job Trotter; with a great Morning of Business in Gray's Inn Square. Concluding with a Double Knock at Mr. Perker's Door
60. Containing some Particulars relative to the Double Knock, and other Matters: among which certain Interesting Disclosures relative to Mr. Snodgrass and a Young Lady are by no means irrelevant to this History
61. Mr. Solomon Pell, assisted by a Select Committee of Coachmen, arranges the affairs of the elder Mr. Weller
62. An important Conference takes place between Mr. Pickwick and Samuel Weller, at which his Parent assists. An old Gentleman in a snuff-coloured Suit arrives unexpectedly
63. In which the Pickwick Club is finally dissolved, and everything concluded to the Satisfaction of Everybody

5. **Oliver Twist**

- 1 PREFACES
 1. PREFACE TO THE THIRD EDITION, 1841
 2. PREFACE TO THE FIRST CHEAP EDITION
 3. PREFACE TO THE 'CHARLES DICKENS' EDITION
- 2 Treats of the Place where Oliver Twist was born, and of the Circumstances attending his Birth
- 3 Treats of Oliver Twist's Growth, Education, and Board
- 4 Relates how Oliver Twist was very near getting a Place which would not have been a Sinecure
- 5 Oliver, being offered another Place, makes his first Entry into Public Life

- 6 Oliver mingles with new Associates. Going to a Funeral for the first Time, he forms an unfavourable Notion of his Master's Business
- 7 Oliver, being goaded by the Taunts of Noah, rouses into Action, and rather astonishes him
- 8 Oliver continues refractory
- 9 Oliver Walks to London. He Encounters on the Road a strange sort of young Gentleman
- 10 Containing further Particulars concerning the Pleasant Old Gentleman, and his hopeful Pupils
- 11 Oliver becomes better acquainted with the Characters of his new Associates; and purchases Experience at a high Price. Being a Short, but very important Chapter, in this History
- 12 Treats of Mr. Fang the Police Magistrate; and furnishes a slight Specimen of his Mode of administering Justice
- 13 In which Oliver is taken better Care of than he ever was before. And in which the Narrative reverts to the Merry Old Gentleman and His youthful Friends.
- 14 Some new Acquaintances are introduced to the intelligent Reader, connected with whom, various pleasant Matters are related, appertaining to this History
- 15 Comprising further Particulars of Oliver's Stay at Mr. Brownlow's, with the remarkable Prediction which one Mr. Grimwig uttered concerning him, when he went out on an Errand
- 16 Showing how very fond of Oliver Twist, the Merry Old Jew and Miss Nancy were
- 17 Relates what Became of Oliver Twist, after he had been claimed by Nancy
- 18 Oliver's Destiny continuing unpropitious, brings a great Man to London to injure his Reputation
- 19 How Oliver passed his Time in the improving Society of his reputable Friends
- 20 In which a notable Plan is discussed and determined on
- 21 Wherein Oliver is delivered over to Mr. William Sikes
- 22 The Expedition
- 23 The Burglary
- 24 Which contains the Substance of a pleasant Conversation between Mr. Bumble and a Lady; and shows that even a Beadle may be susceptible on some Points
- 25 Treats on a very poor Subject. But is a short one, and may be found of Importance in this History
- 26 Wherein this History reverts to Mr. Fagin and Company
- 27 In which a mysterious Character appears upon the Scene; and many Things, inseparable from this History, are done and performed
- 28 Atones for the Unpoliteness of a former Chapter; which deserted a Lady, most unceremoniously
- 29 Looks after Oliver, and proceeds with his Adventures
- 30 Has an introductory Account of the Inmates of the House, to which Oliver resorted
- 31 Relates what Oliver's new visitors thought of him
- 32 Involves a critical Position

- 33 Of the happy Life Oliver began to lead with his kind Friends
- 34 Wherein the happiness of Oliver and His Friends, experiences a sudden Check
- 35 Contains some introductory Particulars relative to a young Gentleman who now arrives upon the Scene; and a new Adventure which happened to Oliver
- 36 Containing the unsatisfactory Result of Oliver's Adventure; and a Conversation of some Importance between Harry Maylie and Rose
- 37 Is a very short one, and may appear of no great Importance in its Place, but it should be read notwithstanding, as a Sequel to the Last, and a Key to one that will follow when its Time arrives
- 38 In which the Reader may perceive a Contrast, not uncommon in matrimonial Cases
- 39 Containing an Account of what passed between Mr. and Mrs. Bumble, and Mr. Monks, at their nocturnal Interview
- 40 Introduces some respectable Characters with whom the Reader is already acquainted, and shows how Monks and the Jew laid their worthy Heads together
- 41 A strange Interview, which is a Sequel to the last Chamber
- 42 Containing fresh Discoveries, and showing that Surprises, like Misfortunes, seldom come alone
- 43 An old Acquaintance of Oliver's, exhibiting decided Marks of Genius, becomes a public Character in the Metropolis
- 44 Wherein is shown how the Artful Dodger got into Trouble
- 45 The Time arrives for Nancy to redeem her Pledge to Rose Maylie. She Fails
- 46 Noah Claypole is employed by Fagin on a secret mission
- 47 The Appointment kept
- 48 Fatal Consequences
- 49 The Flight of Sikes
- 50 Monks and Mr. Brownlow at length meet. Their Conversation, and the Intelligence that interrupts it
- 51 The Pursuit and Escape
- 52 Affording an Explanation of more mysteries than one, and comprehending a proposal of Marriage with no Word of Settlement or Pin-Money
- 53 Fagin's last Night alive
- 54 And Last

6. **Nicholas Nickleby I**

7. PREFACES, ETC.
8. NICKLEBY PROCLAMATION BY 'BOZ' 1838
9. PREFACE TO THE FIRST EDITION
10. PREFACE TO THE FIRST CHEAP EDITION
11. Introduces all the rest
12. Of Mr. Ralph Nickleby, and his Establishments, and his Undertakings, and of a Great Joint Stock Company of Vast National Importance
13. Mr. Ralph Nickleby receives sad Tidings of his Brother, but bears up nobly against the Intelligence communicated to him. The Reader is informed how

- he liked Nicholas, who is herein introduced, and how kindly he proposed to make his Fortune at once
14. Nicholas and his Uncle (to secure the Fortune without loss of time) wait upon Mr. Wackford Squeers, the Yorkshire Schoolmaster
 15. Nicholas starts for Yorkshire. Of his Leave-taking and his Fellow-Travelers, and what befell them on the Road
 16. In which the Occurrence of the Accident mentioned in the last Chapter, affords an Opportunity to a Couple of Gentlemen to tell Stories against each other
 17. Mr. and Mrs. Squeers at Home
 18. Of the Internal Economy of Dotheboys Hall
 19. Of Miss Squeers, Mrs. Squeers, Master Squeers, and Mr. Squeers; and of Various Matters and Persons connected no less with the Squeerses than Nicholas Nickleby
 20. How Mr. Ralph Nickleby provided for his Niece and Sister-in-Law
 21. Newman Noggs inducts Mrs. and Miss Nickleby into their New Dwelling in the City
 22. Whereby the Reader will be enabled to trace the further course of Miss Fanny Squeer's Love, and to ascertain whether it ran smooth or otherwise
 23. Nicholas varies the Monotony of Dotheboys Hall by a most Vigorous and remarkable Proceeding, which leads to Consequences of some Importance
 24. Having the Misfortune to treat of none but Common People, is necessarily of a Mean and Vulgar Character
 25. Acquaints the Reader with the Cause and Origin of the Interruption described in the last Chapter, and with some other Matters necessary to be known
 26. Nicholas seeks to employ himself in a new Capacity, and being unsuccessful, accepts an Engagement as Tutor in a Private Family
 27. Follows the Fortunes of Miss Nickleby
 28. Miss Knag, after doting on Kate Nickleby for three whole Days, makes up her Mind to hate her for evermore. The Causes which led Miss Knag to form this Resolution
 29. Descriptive of a Dinner at Mr. Ralph Nickleby's, and of the Manner in which the Company entertained themselves, before Dinner, at Dinner, and after Dinner
 30. Wherein Nicholas at length encounters his Uncle, to whom he expresses his Sentiments with much Candour. His Resolution
 31. Madam Mantalini finds herself in a Situation of some Difficulty, and Miss Nickleby finds herself in no Situation at all
 32. Nicholas, accompanied by Smike, sallies forth to seek his Fortune. He encounters Mr. Vincent Crummles; and who he was, is herein made manifest
 33. Treats of the Company of Mr. Vincent Crummles, and of his Affairs, Domestic and Theatrical
 34. Of the Great Bespeak for Miss Snevellicci, and the first Appearance of Nicholas upon any Stage

35. Concerning a young Lady from London, who joins the Company, and an elderly Admirer who follows in her Train; with an affecting Ceremony consequent on their Arrival
36. Is fraught with some Danger to Miss Nickleby's Peace of Mind
37. Mrs. Nickleby becomes acquainted with Messrs Pyke and Pluck, whose Affection and Interest are beyond all Bounds
38. Miss Nickleby, rendered desperate by the Persecution of Sir Mulberry Hawk, and the Complicated Difficulties and Distresses which surround her, appeals, as a last resource, to her Uncle for Protection
39. Of the Proceedings of Nicholas, and certain Internal Divisions in the Company of Mr. Vincent Crummles
40. Festivities are held in honour of Nicholas, who suddenly withdraws himself from the Society of Mr. Vincent Crummles and his Theatrical Companions
41. Of Ralph Nickleby and Newman Noggs, and some wise Precautions, the Success or Failure of which will appear in the Sequel
42. Relating chiefly to some remarkable Conversation, and some remarkable Proceedings to which it gives rise

7. Nicholas Nickleby II

43. In which Mr. Ralph Nickleby is relieved, by a very expeditious Process, from all Commerce with his Relations
44. Wherein Mr. Ralph Nickleby is visited by Persons with whom the Reader has been already made acquainted
45. Smike becomes known to Mrs. Nickleby and Kate. Nicholas also meets with new Acquaintances. Brighter Days seem to dawn upon the Family
46. Private and confidential; relating to Family Matters, showing how Mr. Kenwigs underwent violent Agitation, and how Mrs. Kenwigs was as well as could be expected
47. Nicholas finds further Favour in the Eyes of the brothers Cheeryble and Mr. Timothy Linkinwater. The brothers give a Banquet on a great annual Occasion. Nicholas, on returning Home from it, receives a mysterious and important Disclosure from the Lips of Mrs. Nickleby
48. Comprises certain Particulars arising out of a Visit of Condolence, which may prove important hereafter. Smike unexpectedly encounters a very old Friend, who invites him to his House, and will take no Denial
49. In which another old Friend encounters Smike, very opportunely and to some Purpose
50. In which Nicholas falls in Love. He employs a Mediator, whose Proceedings are crowned with unexpected Success, excepting in one solitary Particular
51. Containing some romantic Passages between Mrs. Nickleby and the Gentleman in the Small-Clothes next Door
52. Illustrative of the convivial Sentiment, that the best of Friends must sometimes part
53. Officiates as a kind of Gentleman Usher, in bringing various People together
54. Mr. Ralph Nickleby cuts an old Acquaintance. It would also appear from the Contents hereof, that a Joke, even between Husband and Wife, may be sometimes carried too far

55. Containing Matter of a surprising Kind
56. Throws some Light upon Nicholas's Love; but whether for Good or Evil the Reader must determine
57. Mr. Ralph Nickleby has some confidential Intercourse with another old Friend. They concert between them a Project, which promises well for both
58. Being for the Benefit of Mr. Vincent Crummles, and positively his last Appearance on this Stage
59. Chronicles the further Proceedings of the Nickleby Family, and the Sequel of the Adventure of the Gentleman in the Small-Clothes
60. Involves a serious Catastrophe
61. The Project of Mr. Ralph Nickleby and his Friend approaching a successful issue, becomes unexpectedly known to another Party, not admitted into their Confidence
62. Nicholas despairs of rescuing Madeline Bray, but plucks up his Spirits again, and determines to attempt it. Domestic Intelligence of the Kenwidges and Lillyvicks
63. Containing the further Progress of the Plot contrived by Mr. Ralph Nickleby and Mr. Arthur Gride
64. The Crisis of the Project and its Result
65. Of Family Matters, Cares, Hopes, Disappointments, and Sorrows
66. Ralph Nickleby, baffled by his Nephew in his late Design, hatches a Scheme of Retaliation which Accident suggests to him, and takes into his Counsels a tried Auxiliary
67. How Ralph Nickleby's Auxiliary went about his Work, and how he prospered with it
68. In which one Scene of this History is closed
69. The Plots begin to fail, and Doubts and Dangers to disturb the Plotter
70. The Dangers thicken, and the Worst is told
71. Wherein Nicholas and his Sister forfeit the good Opinion of all worldly and prudent People
72. Ralph makes one last Appointment – and keeps it
73. The Brothers Cheeryble make various Declarations for themselves and others. Tim Linkinwater makes a Declaration for himself
74. An old Acquaintance is recognised under melancholy Circumstances, and Dotheboys Hall breaks up forever
75. Conclusion

8. The Old Curiosity Shop VOL. I

1. PREFACE
2. Chapter the First
3. Chapter the Second
4. Chapter the Third
5. Chapter the forth
6. Chapter the Fifth
7. Chapter the sixth
8. Chapter the Seventh
9. Chapter the Eighth

10. Chapter the Ninth
11. Chapter the Tenth
12. Chapter the Eleventh
13. Chapter the Twelfth
14. Chapter the Thirteenth
15. Chapter the Fourteenth
16. Chapter the Fifteenth
17. Chapter the Sixteenth
18. Chapter the Seventeenth
19. Chapter the Eighteenth
20. Chapter the Nineteenth
21. Chapter the Twentieth
22. Chapter the Twenty-First
23. Chapter the Twenty-Second
24. Chapter the Twenty-Third
25. Chapter the Twenty-Fourth
26. Chapter the Twenty-Fifth
27. Chapter the Twenty-Sixth
28. Chapter the Twenty-Seventh
29. Chapter the Twenty-Eighth
30. Chapter the Twenty-Ninth
31. Chapter the Thirtieth
32. Chapter the Thirty-First
33. Chapter the Thirty-Second
34. Chapter the Thirty-Third
35. Chapter the Thirty-Fourth
36. Chapter the Thirty-Fifth
37. Chapter the Thirty-Sixth
38. Chapter the Thirty-Seventh

9. The Old Curiosity Shop VOL. II

39. Chapter the Thirty-Eighth
40. Chapter the Thirty-Ninth
41. Chapter the fortieth
42. Chapter the forty-First
43. Chapter the forty-Second
44. Chapter the forty-Third
45. Chapter the forty-Fourth
46. Chapter the forty-Fifth
47. Chapter the forty-sixth
48. Chapter the forty-Seventh
49. Chapter the forty-Eight
50. Chapter the forty-Ninth
51. Chapter the Fiftieth
52. Chapter the Fifty-First
53. Chapter the Fifty-Second
54. Chapter the Fifty-Third

55. Chapter the Fifty-Fourth
56. Chapter the Fifty-Fifth
57. Chapter the Fifty-Sixth
58. Chapter the Fifty-Seventh
59. Chapter the Fifty-Eight
60. Chapter the Fifty-Ninth
61. Chapter the Sixtieth
62. Chapter the Sixty-First
63. Chapter the Sixty-Second
64. Chapter the Sixty-Third
65. Chapter the Sixty-Fourth
66. Chapter the Sixty-Fifth
67. Chapter the Sixty-Sixth
68. Chapter the Sixty-Seventh
69. Chapter the Sixty-Eighth
70. Chapter the Sixty-Ninth
71. Chapter the Seventieth
72. Chapter the Seventy-First
73. Chapter the Seventy-Second
74. Chapter the Last

10. Barnaby Rudge VOL. I

1. PREFACE
 1. PREFACE OT THE THIRD VOLUME OF MASTER HUMPHREY'S
CLOCK ('BARNABY RUDGE')
 2. PREFACE TO THE FIRST CHEAP EDITION OF 'BARNABY RUDGE'
2. Chapter the First
3. Chapter the Second
4. Chapter the Third
5. Chapter the forth
6. Chapter the Fifth
7. Chapter the sixth
8. Chapter the Seventh
9. Chapter the Eighth
10. Chapter the Ninth
11. Chapter the Tenth
12. Chapter the Eleventh
13. Chapter the Twelfth
14. Chapter the Thirteenth
15. Chapter the Fourteenth
16. Chapter the Fifteenth
17. Chapter the Sixteenth
18. Chapter the Seventeenth
19. Chapter the Eighteenth
20. Chapter the Nineteenth
21. Chapter the Twentieth
22. Chapter the Twenty-First

23. Chapter the Twenty-Second
24. Chapter the Twenty-Third
25. Chapter the Twenty-Fourth
26. Chapter the Twenty-Fifth
27. Chapter the Twenty-Sixth
28. Chapter the Twenty-Seventh
29. Chapter the Twenty-Eighth
30. Chapter the Twenty-Ninth
31. Chapter the Thirtieth
32. Chapter the Thirty-First
33. Chapter the Thirty-Second
34. Chapter the Thirty-Third
35. Chapter the Thirty-Fourth
36. Chapter the Thirty-Fifth
37. Chapter the Thirty-Sixth
38. Chapter the Thirty-Seventh
39. Chapter the Thirty-Eighth
40. Chapter the Thirty-Ninth
41. Chapter the fortieth
42. Chapter the forty-First
43. Chapter the forty-Second

11. Barnaby Rudge VOL. II

44. Chapter the forty-Third
45. Chapter the forty-Fourth
46. Chapter the forty-Fifth
47. Chapter the forty-Sixth
48. Chapter the forty-Seventh
49. Chapter the forty-Eight
50. Chapter the forty-Ninth
51. Chapter the Fiftieth
52. Chapter the Fifty-First
53. Chapter the Fifty-Second
54. Chapter the Fifty-Third
55. Chapter the Fifty-Fourth
56. Chapter the Fifty-Fifth
57. Chapter the Fifty-Sixth
58. Chapter the Fifty-Seventh
59. Chapter the Fifty-Eight
60. Chapter the Fifty-Ninth
61. Chapter the Sixtieth
62. Chapter the Sixty-First
63. Chapter the Sixty-Second
64. Chapter the Sixty-Third
65. Chapter the Sixty-Fourth
66. Chapter the Sixty-Fifth
67. Chapter the Sixty-Sixth

68. Chapter the Sixty-Seventh
69. Chapter the Sixty-Eighth
70. Chapter the Sixty-Ninth
71. Chapter the Seventieth
72. Chapter the Seventy-First
73. Chapter the Seventy-Second
74. Chapter the Seventy-Third
75. Chapter the Seventy-fourth
76. Chapter the Seventy-fifth
77. Chapter the Seventy-sixth
78. Chapter the Seventy-seventh
79. Chapter the Seventy-eighth
80. Chapter the Seventy-ninth
81. Chapter the Eightieth
82. Chapter the Eighty-first
83. Chapter the Last

12. American Notes for General Circulation – Pictures from Italy

1. AMERICAN NOTES

1. Dedication Page
2. PREFACES
3. Preface to the First Cheap Edition
4. Preface to the 'Charles Dickens' Edition
5. Chapter I. Going Away
6. Chapter II. The Passage Out
7. Chapter III. Boston
8. Chapter IV. An American Railroad. Lowell and its Factory System
9. Chapter V. Worcester. The Connecticut River. Hartford. New Haven. To New York
10. Chapter VI. New York
11. Chapter VII. Philadelphia, and its Solitary Prison
12. Chapter VIII. Washington. The Legislature. And the President's House
13. Chapter IX. A Night Steamer on the Potomac River. Virginia Road, and a Black Driver. Richmond. Baltimore. The Harrisburg Mail, and a Glimpse of the City. A Canal Boat
14. Chapter X. Some further Account of the Canal Boat, its Domestic Economy, and its Passengers. Journey to Pittsburg across the Alleghany Mountains. Pittsburg
15. Chapter XI. From Pittsburg to Cincinnati in a Western Steamboat. Cincinnati
16. Chapter XII. From Cincinnati to Louisville in another Western Steamboat; and from Louisville to St. Louis in Another. St. Louis
17. Chapter XIII. A Jaunt to the Looking-glass Prairie and back
18. Chapter XIV. Return to Cincinnati. A Stage-Coach Ride from that City to Columbus, and thence to Sandusky. So, by Lake Erie, to the Falls of Niagara

19. Chapter XV. In Canada; Toronto; Kingston; Montreal; Quebec; St. John's. In the United States again; Lebanon; the Shaker Village; West Point
20. Chapter XVI. The Passage Home
21. Chapter XVII. Slavery
22. Chapter XVIII. Concluding Remarks
23. Postscript

2. PICTURES FROM ITALY

1. The Reader's Passport
2. Going through France
3. Lyons, the Rhone, and the Goblin of Avignon
4. Avignon to Genoa
5. Genoa and its Neighbourhood
6. To Parma, Modena, and Bologna
7. Through Bologna and Ferrara
8. An Italian Dream
9. By Verona, Mantua, and Milan, across the Pass of the Simplon into Switzerland
10. To Rome by Pisa and Siena
11. Rome
12. A Rapid Diorama –
 1. To Naples
 2. Naples
 3. Pompeii—Herculaneum
 4. Paestum
 5. Vesuvius
 6. Return to Naples
 7. Monte Cassino
 8. Florence

13. A Child's History of England

1. Dedication Page
2. Chapter I. Ancient England and the Romans. From 50 years before Christ, to the year of our Lord 450.
3. Chapter II. Ancient England under the Early Saxons. From the year 450, to the year 871.
4. Chapter III. England under the Good Saxon, Alfred, and Edward the Elder. From the year 871, to the year 901.
5. Chapter IV. England under Athelstan and the Six Boy-Kings. From the year 925, to the year 1016.
6. Chapter V. England under Canute the Dane. From the year 1016, to the year 1035.
7. Chapter VI. England under Harold Harefoot, Hardicanute, and Edward the Confessor. From the year 1035, to the year 1066.
8. Chapter VII. England under Harold the Second, and Conquered by the Normans. All in the same year, 1066.

9. Chapter VIII. England under William the First, the Norman Conqueror. From the year 1066, to the year 1087.
10. Chapter IX. England under William the Second, called Rufus. From the year 1087, to the year 1100.
11. Chapter X. England under Henry the First, called Fine-Scholar. From the year 1100, to the year 1135.
12. Chapter XI. England under Matilda and Stephen. From the year 1135, to the year 1154.
13. Chapter XII. England under Henry the Second. Parts First and Second. From the year 1154, to the year 1189.
14. Chapter XIII. England under Richard the First, called the Lion-Heart. From the year 1189, to the year 1199.
15. Chapter XIV. England under John, called Lackland. From the year 1199, to the year 1216.
16. Chapter XV. England under Henry the Third, called, of Winchester. From the year 1216, to the year 1272.
17. Chapter XVI. England under Edward the First, called Longshanks. From the year 1172, to the year 1307.
18. Chapter XVII. England under Edward the Second. From the year 1307, to the year 1327.
19. Chapter XVIII. England under Edward the Third. From the year 1327, to the year 1377.
20. Chapter XIX. England under Richard the Second. From the year 1377, to the year 1399.
21. Chapter XX. England under Henry the Fourth, called Bolingbroke. From the year 1399, to the year 1413.
22. Chapter XXI. England under Henry the Fifth. Parts First and Second. From the year 1413, to the year 1422.
23. Chapter XXII. England under Henry the Sixth. Parts First and Second (The Story of Joan of Arc), and Third. From the year 1422, to the year 1461.
24. Chapter XXIII. England under Edward the Fourth. From the year 1461, to the year 1483.
25. Chapter XXIV. England under Edward the Fifth. For few weeks in the year 1483.
26. Chapter XXV. England under Richard the Third. From the year 1483, to the year 1485.
27. Chapter XXVI. England under Henry the Seventh. From the year 1485, to the year 1509.
28. Chapter XXVII. England under Henry the Eighth, called Bluff King Hal and Burly King Harry. From the year 1509, to the year 1533.
29. Chapter XXVIII. England under Henry the Eighth, called Bluff King Hal and Burly King Harry. From the year 1533, to the year 1547.
30. Chapter XXIX. England under Edward the Sixth. From the year 1547, to the year 1553.
31. Chapter XXX. England under Mary. From the year 1553, to the year 1558.
32. Chapter XXXI. England under Elizabeth. Parts First, Second, and Third. From the year 1553, to the year 1603.

33. Chapter XXXII. England under James the First. Parts First and Second. From the year 1603, to the year 1625.
34. Chapter XXXIII. England under Charles the First. Parts First, Second, Third and Fourth. From the year 1625, to the year 1649.
35. Chapter XXXIV. England under Oliver Cromwell. Parts First and Second. From the year 1649, to the year 1660.
36. Chapter XXXV. England under Charles the Second, called the Merry Monarch. Parts First and Second. From the year 1660, to the year 1685.
37. Chapter XXXVI. England under James the Second. From the year 1685, to the year 1688.
38. Chapter XXXVII. Conclusion. From the year 1688, to the year 1837.

14. Martin Chuzzlewit VOL. I

1. PREFACES
 1. PREFACE TO THE FIRST EDITION
 2. PREFACE TO THE FIRST CHEAP EDITION
 3. PREFACE TO THE 'CHARLES DICKENS' EDITION
2. Introductory, concerning the Pedigree of the Chuzzlewit Family
3. Wherein certain Persons are presented to the Reader, with whom he may, if he please, become better acquainted
4. In which certain other Persons are introduced; on the same Terms as in the last Chapter
5. From which it will appear that if Union be Strength, and Family Affection be pleasant to contemplate, the Chuzzlewits were the strongest and most agreeable Family in the World
6. Containing a full Account of the Installation of Mr. Pecksniff's new Pupil into the Bosom of Mr. Pecksniff's Family. With all the Festivities held on that Occasion, and the great Enjoyment of Mr. Pinch
7. Comprises, among other important matters, Pecksniffian and Architectural, an exact Relation of the Progress made by Mr. Pinch in the Confidence and Friendship of the new Pupil
8. In Which Mr. Chevy Slyme asserts the Independence of his Spirit, and the Blue Dragon Loses a Limb
9. Accompanies Mr. Pecksniff and his charming Daughters to the City of London; and relates what fell out upon their way thither
10. Town and Todgers's
11. Containing strange Matter, on which many Events in this History may, for their good or evil Influence, chiefly depend
12. Wherein a certain Gentleman becomes particular in his Attentions to a certain Lady; and more Coming Events than one, cast their Shadows before
13. Will be seen in the Long Run, if not in the Short One, to concern Mr. Pinch and others, nearly. Mr. Pecksniff asserts the Dignity of outraged Virtue. Young Martin Chuzzlewit forms a desperate Resolution
14. Showing what became of Martin and His desperate Resolve, after he left Mr. Pecksniff's House; what Persons he encountered; what Anxieties he Suffered; and what News he heard

15. In which Martin bids Adieu to the Lady of his Love; and honours an obscure Individual whose Fortune he intends to make by commending her to his protection
16. The Burden whereof, is Hail Columbia!
17. Martin disembarks from that noble and fast-sailing Line-of-Packet Ship, the Screw, at the Port of New York, in the United States of America. He makes some Acquaintances, and dines at a Boarding-House. The Particulars of those Transactions
18. Martin enlarges his Circle of acquaintance; increases his Stock of Wisdom; and has an excellent Opportunity of comparing his own Experiences with those of Lummy Ned of the Light Salisbury, as related by his Friend Mr. William Simmons
19. Does Business with the House of Anthony Chuzzlewit and Son, from which one of the Partners retires unexpectedly
20. The Reader is brought into Communication with some Professional Persons, and sheds a tear over the Filial Piety of good Mr. Jonas
21. Is a Chapter of Love
22. More American Experiences, Martin takes a Partner, and makes a Purchase. Some Account of Eden, as it appeared on Paper. Also of the British Lion. Also of the Kind of Sympathy professed and entertained by the Watertoast Association of United Sympathisers
23. From which it will be seen that Martin became a Lion of his own Account. Together with the Reason why
24. Martin and his Partner take Possession of their Estate. The joyful occasion involves some further Account of Eden
25. Reports Progress in certain homely Matters of Love, Hatred, Jealousy, and Revenge
26. Is in part professional; and furnishes the Reader with some valuable hints in relation to the Management of a Sick Chamber

15. Martin Chuzzlewit VOL. II

27. An unexpected Meeting, and a promising Prospect
28. Showing that old Friends may not only appear with new Faces, but in false Colours. That People are prone to bite, and that Biters may sometimes be bitten.
29. Mr. Montague at Home. And Mr. Jonas Chuzzlewit at Home
30. In which some People are precocious, others professional, and others mysterious: all in their several ways
31. Proves that changes may be rung in the best-regulated Families, and that Mr. Pecksniff was a special hand at a Triple-Bob-Major
32. Mr. Pinch is discharged of a Duty which he never owed to anybody; and Mr. Pecksniff discharges a Duty which he owes to Society
33. Treats of Todgers's again; and of another blighted Plant besides the Plants upon the Leads
34. Further proceedings in Eden, and a proceeding out of it. Martin makes a Discovery of some Importance

35. In Which the Travellers move Homeward, and encounter some distinguished Characters upon the way
36. Arriving in England, Martin witnesses a Ceremony, from which he derives the cheering Information that he has not been forgotten in his Absence
37. Tom Pinch departs to seek his Fortune. What he finds at starting
38. Tom Pinch, going astray, finds that he is not the only Person in that Predicament. He retaliates upon a fallen Foe
39. Secret Service
40. Containing some further Particulars of the Domestic Economy of the Pinches; with strange News from the City, narrowly concerning Tom
41. The Pinches make a new Acquaintance, and have fresh occasion for Surprise and Wonder
42. Mr. Jonas and his Friend, arriving at a pleasant Understanding, set forth upon an Enterprise
43. Continuation of the Enterprise of Mr. Jonas and his Friend
44. Has an influence on the fortunes of several People. Mr. Pecksniff is exhibited in the Plenitude of Power, and wields the same with Fortitude and Magnanimity
45. Further Continuation of the Enterprise of Mr. Jonas and his Friend
46. In Which Tom Pinch and his Sister take a little Pleasure; but quite in a domestic way, and with no Ceremony about it
47. In Which Miss Pecksniff makes Love, Mr. Jonas makes Wrath, Mrs. Gamp makes Tea, and Mr. Chuffey makes Business
48. Conclusion of the Enterprise of Mr. Jonas and his Friend
49. Bears Tidings of Martin and of Mark, as well as of a third Person not quite unknown to the Reader. Exhibits Filial Piety in an ugly Aspect; and casts a doubtful Ray of Light upon a very dark Place
50. In Which Mrs. Harris assisted by a Tea-pot, is the Cause of a Division between Friends
51. Surprises Tom Pinch very much, and shows how certain Confidences passed between him and his Sister
52. Sheds new and brighter Light upon the very dark Place; and contains the Sequel of the Enterprise of Mr. Jonas and his Friend
53. In which the Tables are turned, completely upside down
54. What John Westlock said to Tom Pinch's Sister; what Tom Pinch's Sister Said to John Westlock; what Tom Pinch said to both of them; and how they all passed the remainder of the day
55. Gives the Author great Concern. For it is the last in the Book
56. POSTSCRIPT

16. Christmas Books

1. PREFACES
 1. PREFACE TO THE FIRST CHEAP EDITION
 2. PREFACE TO THE 'CHARLES DICKENS' EDITION
2. A Christmas Carol in Prose: Being a Short Story of Christmas
 1. Stave I. Marley's Ghost
 2. Stave II. The First of the Three Spirits

3. Stave III. The Second of the Three Spirits
4. Stave IV. The Last of the Spirits
5. Stave V. the End of It
3. The Chimes: A Goblin Story of Some Bells that rang an Old Year out and a New Year in.
 1. First Quarter.
 2. The Second Quarter.
 3. Third Quarter.
 4. Fourth Quarter.
4. The Cricket on the Hearth
 1. Chirp the First
 2. Chirp the Second
 3. Chirp the Third
5. The Battle of Life: A Love Story
 1. Part the First
 2. Part the Second
 3. Part the Third
6. The Haunted Man and the Ghost's Bargain: A Fancy for Christmas Time
 1. The Gift Bestowed
 2. The Gift Diffused
 3. The Gift Reversed

17. Hard Times and Other Stories

1. Book the First—Sowing
 1. The One Thing Needful
 2. Murdering the Innocents
 3. A Loophole
 4. Mr. Bounderby
 5. The Key-Note
 6. Sleary's Horsemanship
 7. Mrs. Sparsit
 8. Never Wonder
 9. Sissy's Progress
 10. Stephen Blackpool
 11. No Way Out
 12. The Old Woman
 13. Rachael
 14. The Great Manufacturer
 15. Father and Daughter
 16. Husband and Wife
2. Book the Second—Reaping
 1. Effects in the Bank
 2. Mr. James Harthouse
 3. The Whelp
 4. Men and Brothers
 5. Men and Masters
 6. Fading Away

7. Gunpowder
8. Explosion
9. Hearing the Last of It
10. Mrs. Sparsit's Staircase
11. Lower and Lower
12. Down
3. Book the Third—Garnering
 1. Another Thing Needful
 2. Very Ridiculous
 3. Very Decided
 4. Lost
 5. Found
 6. The Starlight
 7. Whelp-Hunting
 8. Philosophical
 9. Final

Hunted Down

1. First Chapter
2. Second Chapter
3. Third Chapter
4. Fourth Chapter
5. Fifth Chapter

A Holiday Romance

1. Part I. Introductory Romance from the Pen of William Tinkling, Esq.
2. Part II. Romance. From the Pen of Miss Alice Rainbird
3. Part III. Romance. From the Pen of Lieut.-Col. Robin Redforth
4. Part IV. Romance. From the Pen of Miss Nettie Ashford.

George Silverman's Explanation [1868]

1. First Chapter
2. Second Chapter
3. Third Chapter
4. Fourth Chapter
5. Fifth Chapter
6. Sixth Chapter
7. Seventh Chapter
8. Eighth Chapter
9. Ninth Chapter

18. Dombey and Son VOL. I

1. PREFACES
 1. PREFACE TO THE FIRST EDITION
 2. PREFACE TO THE 'CHARLES DICKENS' EDITION
2. Dombey and Son
3. In which timely Provision is made for an Emergency that will sometimes arise in the best-regulated Families
4. In which Mr. Dombey, as a Man and a Father, is seen at the Head of the Home Department

5. In which some more First Appearances are made on the Stage of these Adventures
6. Paul's Progress and Christening
7. Paul's Second Deprivation
8. A Bird's-eye Glimpse of Miss Tox's Dwelling-place: also of the state of Miss Tox's Affections
9. Paul's further Progress, Growth and Character
10. In which the Wooden Midshipman gets into Trouble
11. Containing the Sequel of the Midshipman's Disaster
12. Paul's Introduction to a New Scene
13. Paul's Education
14. Shipping Intelligence and Office Business
15. Paul grows more and more old-fashioned, and goes Home for the Holidays
16. Amazing Artfulness of Captain Cuttle, and a new Pursuit for Walter Gay
17. What the Waves were always saying
18. Captain Cuttle does a little Business for the Young People
19. Father and Daughter
20. Walter goes away
21. Mr. Dombey goes upon a Journey
22. New Faces
23. A Trifle of Management by Mr. Carker the Manager
24. Florence solitary, and the Midshipman mysterious
25. The Study of a Loving Heart
26. Strange News of Uncle Sol
27. Shadows of the Past and Future
28. Deeper Shadows
29. Alterations
30. The opening of the Eyes of Mrs. Chick
31. The Interval before the Marriage

19. Dombey and Son VOL. II

32. The Wedding
33. The Wooden Midshipman goes to Pieces
34. Contrasts
35. Another Mother and Daughter
36. The Happy Pair
37. Housewarming
38. More Warnings than One
39. Miss Tox improves an Old Acquaintance
40. Further Adventures of Captain Edward Cuttle, Mariner
41. Domestic Relations
42. New Voices in the Waves
43. Confidential and Accidental
44. The Watches of the Night
45. A Separation
46. The Trusty Agent
47. Recognisant and Reflective

48. The Thunderbolt
49. The Flight of Florence
50. The Midshipman makes a Discovery
51. Mr. Toots's Complaint
52. Mr. Dombey and the World
53. Secret Intelligence
54. More Intelligence
55. The Fugitives
56. Rob the Grinder loses his Place
57. Several People delighted, and the Game Chicken disgusted
58. Another Wedding
59. After a Lapse
60. Retribution
61. Chiefly Matrimonial
62. Relenting
63. Final

20. David Copperfield VOL. I

1. PREFACES
 1. PREFACE TO THE FIRST EDITION
 2. PREFACE TO THE 'CHARLES DICKENS' EDITION
2. I am Born
3. I Observe
4. I have a Change
5. I fall into Disgrace
6. I am sent away from Home
7. I enlarge my Circle of Acquaintance
8. My 'First Half' at Salem House
9. My Holidays. Especially one happy Afternoon
10. I have a memorable Birthday
11. I become neglected, and am provided for
12. I begin Life on my own Account, and don't Like it
13. Liking Life on my own Account no better, I form a great Resolution
14. The Sequel of my Resolution
15. My Aunt makes up her Mind about me
16. I make another Beginning
17. I am a New Boy in More Senses than One
18. Somebody turns up
19. A Retrospect
20. I look about me, and make a Discovery
21. Steerforth's Home
22. Little Em'ly
23. Some old Scenes, and some new People
24. I corroborate Mr. Dick, and choose a Profession
25. My first Dissipation
26. Good and bad Angels
27. I fall into Captivity

28. Tommy Traddles
29. Mr. Micawber's Gauntlet
30. I visit Steerforth at his Home, again

21. David Copperfield VOL. II

31. A Loss
32. A greater Loss
33. The beginning of a long Journey
34. Blissful
35. My Aunt astonishes me
36. Depression
37. Enthusiasm
38. A Little Cold Water
39. A Dissolution of Partnership
40. Wickfield and Heep
41. The Wanderer
42. Dora's Aunts
43. Mischief
44. Another Retrospect
45. Our Housekeeping
46. Mr. Dick fulfils my Aunt's Predictions
47. Intelligence
48. Martha
49. Domestic
50. I am involved in Mystery
51. Mr. Peggotty's Dream comes true
52. The Beginning of a Longer Journey
53. I assist at an Explosion
54. Another Retrospect
55. Mr. Micawber's Transactions
56. Tempest
57. The New Wound, and the old
58. The Emigrants
59. Absence
60. Return
61. Agnes
62. I am shown two interesting Penitents
63. A Light shines on my Way
64. A Visitor
65. A last Retrospect

22. Bleak House VOL. I

1. PREFACES
 1. PREFACE TO THE FIRST EDITION
 2. PREFACE TO THE 'CHARLES DICKENS' EDITION
2. In Chancery
3. In Fashion

4. A Progress
5. Telescopic Philanthropy
6. A Morning Adventure
7. Quite at Home
8. The Ghost's Walk
9. Covering a Multitude of Sins
10. Signs and Tokens
11. The Law-Writer
12. Our dear Brother
13. On the Watch
14. Esther's Narrative
15. Deportment
16. Bell Yard
17. Tom-all-Alone's
18. Esther's Narrative
19. Lady Dedlock
20. Moving On
21. A new Lodger
22. The Smallweed Family
23. Mr. Bucket
24. Esther's Narrative
25. An Appeal Case
26. Mrs. Snagsby sees it all
27. Sharpshooters
28. More old Soldiers than one
29. The Ironmaster
30. The Young Man
31. Esther's Narrative
32. Nurse and Patient

23. Bleak House VOL. II

33. The Appointed Time
34. Interlopers
35. A Turn of the Screw
36. Esther's Narrative
37. Chesney Wold
38. Jarndyce and Jarndyce
39. A Struggle
40. Attorney and Client
41. National and Domestic
42. In Mr. Tulkinghorn's Room
43. In Mr. Tulkinghorn's Chambers
44. Esther's Narrative
45. The Letter and the Answer
46. In Trust
47. Stop him!
48. Jo's Will

49. Closing in
50. Dutiful Friendship
51. Esther's Narrative
52. Enlightened
53. Obstinacy
54. The Track
55. Springing a Mine
56. Flight
57. Pursuit
58. Esther's Narrative
59. A Wintry Day and Night
60. Esther's Narrative
61. Perspective
62. A Discovery
63. Another Discovery
64. Steel and Iron
65. Esther's Narrative
66. Beginning the World
67. Down in Lincolnshire
68. The Close of Esther's Narrative

24. Little Dorrit VOL. I

1. PREFACES
 1. PREFACE TO THE FIRST EDITION
 2. PREFACE TO THE 'CHARLES DICKENS' EDITION
2. Book the First: Poverty
3. Sun and Shadow
4. Fellow Travellers
5. Home
6. Mrs. Flintwinch has a Dream
7. Family Affairs
8. The Father of the Marshalsea
9. The Child of the Marshalsea
10. The Lock
11. Little Mother
12. Containing the whole Science of Government
13. Let Loose
14. Bleeding Heart Yard
15. Patriarchal
16. Little Dorrit's Party
17. Mrs. Flintwinch has another Dream
18. Nobody's Weakness
19. Nobody's Rival
20. Little Dorrit's Lover
21. The Father of the Marshalsea in two or three Relations
22. Moving in Society
23. Mr. Merdle's Complaint

24. A Puzzle
25. Machinery in Motion
26. Fortune-Telling
27. Conspirators and Others
28. Nobody's State of Mind
29. Five-and-Twenty
30. Nobody's Disappearance
31. Mrs. Flintwinch goes on dreaming
32. The Word of a Gentleman
33. Spirit
34. More for tune-Telling
35. Mrs. Merdle's Complaint
36. A Shoal of Barnacles
37. What was behind Mr. Pancks on Little Dorrit's Hand
38. The Marshalsea becomes an Orphan

25. Little Dorrit VOL. II. Book the Second—Riches

39. Fellow-Travellers
40. Mrs. General
41. On the Road
42. A Letter from Little Dorrit
43. Something Wrong Somewhere
44. Something Right Somewhere
45. Mostly, Prunes and Prism
46. The Dowager Mrs. Gowan is reminded that it never does
47. Appearance and Disappearance
48. The Dreams of Mrs. Flintwinch thicken
49. A Letter from Little Dorrit
50. In which a great Patriotic Conference is holden
51. The Progress of an Epidemic
52. Taking Advice
53. No just Cause or Impediment why these two Persons should not be joined together
54. Getting on
55. Missing
56. A Castle in the Air
57. The Storming of the Castle in the Air
58. Introduces the next
59. The History of a Self Tormentor
60. Who passes by this Road so late
61. Mistress Affery makes a Conditional Promise, respecting her Dreams
62. The Evening of a Long Day
63. The Chief Butler resigns the Seals of Office
64. Reaping the Whirlwind
65. The Pupil of the Marshalsea
66. An Appearance in the Marshalsea
67. A Plea in the Marshalsea

- 68. Closing in
- 69. Closed
- 70. Going
- 71. Going!
- 72. Gone

26. Christmas Stories from 'Household Words' and 'All the Year Round' VOL. I

- 1. FROM HOUSEHOLD WORDS
- 2. A Christmas Tree
- 3. What Christmas is as we Grow Older
- 4. The Poor Relation's Story
- 5. The Child's Story
- 6. The Schoolboy's Story
- 7. Nobody's Story
- 8. The Seven Poor Travellers
- 9. The Holly-Tree
- 10. Wreck of the Golden Mary [with Wilkie Collins 1856]
- 11. The Perils of Certain English Prisoners [with Wilkie Collins]
- 12. Going into Society
- 13. FROM 'ALL THE YEAR AROUND'
- 14. The Haunted House
- 15. A Message from the Sea
- 16. Tom Tiddler's Ground
- 17. Somebody's Luggage

27. Christmas Stories from 'Household Words' and 'All the Year Round' VOL. II

- 18. Mrs. Lirriper's Lodgings
- 19. Mrs. Lirriper's Legacy
- 20. Doctor Marigold
- 21. Mugby Junction
- 22. No Thoroughfare [with Wilkie Collins]
- 23. The Lazy Tour of Two Idle Apprentices [with Wilkie Collins]

28. A Tale of Two Cities

- 1. BOOK THE FIRST—RECALLED TO LIFE
 - 1. The Period
 - 2. The Mail
 - 3. The Night Shadows
 - 4. The Preparation
 - 5. The Wine-shop
 - 6. The Shoemaker
- 2. BOOK THE SECOND—THE GOLDEN THREAD
 - 1. Five Years Later
 - 2. A Sight
 - 3. A Disappointment
 - 4. Congratulatory

5. The Jackal
6. Hundreds of People
7. Monseigneur in Town
8. Monseigneur in the Country
9. The Gorgon's Head
10. Two Promises
11. A Companion Picture
12. The Fellow of Delicacy
13. The Fellow of no Delicacy
14. The Honest Tradesman
15. Knitting
16. Still Knitting
17. One Night
18. Nine Days
19. An Opinion
20. A Plea
21. Echoing Footsteps
22. The Sea still rises
23. Fire rises
24. Drawn to the Loadstone Rock
3. BOOK THE THIRD—THE TRACK OF A STORM
 1. In Secret
 2. The Grindstone
 3. The Shadow
 4. Calm in Storm
 5. The Wood-Sawyer
 6. Triumph
 7. A Knock at the Door
 8. A Hand at Cards
 9. The Game made
 10. The Substance of the Shadow
 11. Dusk
 12. Darkness
 13. Fifty-Two
 14. The Knitting done
 15. The Footsteps die out for ever

29. Great Expectations

1. Chapter the First
2. Chapter the Second
3. Chapter the Third
4. Chapter the Forth
5. Chapter the Fifth
6. Chapter the Sixth
7. Chapter the Seventh
8. Chapter the Eighth
9. Chapter the Ninth

10. Chapter the Tenth
11. Chapter the Eleventh
12. Chapter the Twelfth
13. Chapter the Thirteenth
14. Chapter the Fourteenth
15. Chapter the Fifteenth
16. Chapter the Sixteenth
17. Chapter the Seventeenth
18. Chapter the Eighteenth
19. Chapter the Nineteenth
20. Chapter the Twentieth
21. Chapter the Twenty-First
22. Chapter the Twenty-Second
23. Chapter the Twenty-Third
24. Chapter the Twenty-Fourth
25. Chapter the Twenty-Fifth
26. Chapter the Twenty-Sixth
27. Chapter the Twenty-Seventh
28. Chapter the Twenty-Eighth
29. Chapter the Twenty-Ninth
30. Chapter the Thirtieth
31. Chapter the Thirty-First
32. Chapter the Thirty-Second
33. Chapter the Thirty-Third
34. Chapter the Thirty-Fourth
35. Chapter the Thirty-Fifth
36. Chapter the Thirty-Sixth
37. Chapter the Thirty-Seventh
38. Chapter the Thirty-Eighth
39. Chapter the Thirty-Ninth
40. Chapter the Fortieth
41. Chapter the Forty-First
42. Chapter the Forty-Second
43. Chapter the Forty-Third
44. Chapter the Forty-Fourth
45. Chapter the Forty-Fifth
46. Chapter the Forty-Sixth
47. Chapter the Forty-Seventh
48. Chapter the Forty-Eight
49. Chapter the Forty-Ninht
50. Chapter the Fiftieth
51. Chapter the Fifty-First
52. Chapter the Fifty-Second
53. Chapter the Fifty-Third
54. Chapter the Fifty-Fourth
55. Chapter the Fifty-Fifth
56. Chapter the Fifty-Sixth

57. Chapter the Fifty-Seventh
58. Chapter the Fifty-Eight
59. Chapter the Fifty-Ninth

30. The Uncommercial Traveller

1. His General Line of Business
2. The Shipwreck
3. Wapping Workhouse
4. Two Views of a Cheap Theatre
5. Poor Mercantile Jack
6. Refreshments for Travellers
7. Travelling Abroad
8. The Great Tasmania's Cargo
9. City of London Churches
10. Shy Neighbourhoods
11. Tramps
12. Dullborough Town
13. Night Walks
14. Chambers
15. Nurse's Stories
16. Arcadian London
17. The Italian Prisoner
18. The Calais Night Mail
19. Some Recollections of Mortality
20. Birthday Celebrations
21. The Short-Timers
22. Bound for the Great Salt Lake
23. The City of the Absent
24. An Old Stage-coaching House
25. The Boiled Beef of New England
26. Chatham Dockyard
27. In the French-Flemish Country
28. Medicine Men of Civilisation
29. Titbull's Alms-Houses
30. The Ruffian
31. Aboard Ship
32. A Small Star in the East
33. A Little Dinner in an Hour
34. Mr. Barlow
35. On An Amateur Beat
36. A Fly-leaf in a Life
37. A Plea for Total Abstinence

31. Our Mutual Friend VOL. I

1. BOOK THE FIRST—THE CUP AND THE LIP
 1. On the Look-out
 2. The Man from Somewhere

3. Another Man
 4. The R. Wilfer Family
 5. Boffin's Bower
 6. Cut adrift
 7. Mr. Wegg looks after himself
 8. Mr. Boffin in Consultation
 9. Mr. and Mrs. Boffin in Consultation
 10. A Marriage Contract
 11. Podsnappery
 12. The Sweat of an honest Man's Brow
 13. Tracking the Bird of Prey
 14. The Bird of Prey brought down
 15. Two new Servants
 16. Minders and Reminders
 17. A Dismal Swamp
2. BOOK THE SECOND: BIRDS OF A FEATHER
1. Of an Educational Character
 2. Still Educational
 3. A Piece of Work
 4. Cupid prompted
 5. Mercury prompting
 6. A Riddle without an Answer
 7. In which a Friendly Move is Originated
 8. In which an Innocent Elopement Occurs
 9. In which the Orphan makes his Will
 10. A Successor
 11. Some Affairs of the Heart
 12. More Birds of Prey
 13. A Solo and a Duett
 14. Strong of Purpose
 15. The whole Case so far
 16. An Anniversary Occasion

32. Our Mutual Friend VOL. II

3. BOOK THE THIRD—A LONG LANE
1. Lodgers in Queer Street
 2. A Respected Friend in a new Aspect
 3. The Same Respected Friend in more Aspects than one
 4. A Happy Return of the Day
 5. The Golden Dustman falls into bad Company
 6. The Golden Dustman falls into worse Company
 7. The Friendly Move takes up a strong Position
 8. The End of a Long Journey
 9. Somebody becomes the Subject of a Prediction
 10. Scouts Out
 11. In the Dark
 12. Meaning Mischief

13. Give a Dog a bad Name, and Hang Him
14. Mr. Wegg Prepares a Grindstone for Mr. Boffin's Nose
15. The Golden Dustman at His Worst
16. The Feast of the Three Hobgoblins
17. A Social Chorus
4. BOOK THE FOURTH—A TURNING
 1. Setting Traps
 2. The Golden Dustman rises a Little
 3. The Golden Dustman sinks again
 4. A Runaway Match
 5. Concerning the Mendicant's Bride
 6. A Cry for Help
 7. Better to be Abel than Cain
 8. A few Grains of Pepper
 9. Two Places vacated
 10. The Dolls' Dressmaker discovers a Word
 11. Effect is given to the Dolls' Dressmaker's Discovery
 12. The Passing Shadow
 13. Showing how the Golden Dustman helped to scatter Dust
 14. Checkmate to the Friendly Move
 15. What was caught in the Traps that were set
 16. Persons and Things in General
 17. CHAPTER THE LAST. The Voice of Society
 18. POSTSCRIPT IN LIEU OF PREFACE
 19. Postscript: In Lieu of Preface

33. The Mystery of Edwin Drood – and Master Humphrey's Clock

1. THE MYSTERY OF EDWIN DROOD
 1. The Dawn
 2. A Dean, and a Chapter also
 3. The Nuns' House
 4. Mr. Sapsea
 5. Mr. Durdles and Friend
 6. Philanthropy in Minor Canon Corner
 7. More Confidences than one
 8. Daggers drawn
 9. Birds in the Bush
 10. Smoothing the Way
 11. A Picture and a Ring
 12. A Night with Durdles
 13. Both at their best
 14. When shall these three meet again?
 15. Impeached
 16. Devoted
 17. Philanthropy, Professional and Unprofessional
 18. A Settler in Cloisterham
 19. Shadow on the Sun-Dial

20. A Flight
21. A Recognition
22. A Gritty State of Things comes on
23. The Dawn again

1. FRAGMENT

1. HOW MR. SAPSEA CEASED TO BE A MEMBER OF THE EIGHT CLUB. TOLD BY HIMSELF

2. MASTER HUMPHREY'S CLOCK

1. DEDICATION
2. PREFACES
3. Master Humphrey, from his Clock-side in the Chimney Corner
4. Master Humphrey, from His Clock-side in the Chimney-Corner
5. Master Humphrey's Visitor
6. The Clock
7. Mr. Weller's Watch
8. Chapter VI
9. Master Humphrey, from His Clock-side in the Chimney Corner

34. Reprinted Pieces and Sunday under Three Heads and Other Tales, Sketches, Articles, etc.

1. REPRINTED PIECES

1. A Child's Dream of a Star
2. The Begging-Letter Writer
3. A Walk in a Workhouse
4. The Ghost of Art
5. The Detective Police
6. Three 'Detective' Anecdotes
7. The Pair of Gloves
8. The Artful Touch
9. The Sofa
10. A Poor Man's Tale of a Patent
11. Births. Mrs. Meek, of a Son (added from the National Edition).
12. A Monument of French Folly
13. Bill-Sticking
14. On Duty with Inspector Field
15. Our English Watering-Place
16. A Flight
17. Our School
18. A Plated Article
19. Our Honourable Friend
20. Our Vestry
21. Our Bore
22. Lying Awake
23. Down with the Tide
24. The Noble Savage
25. The Long Voyage
26. Our French Watering-Place

27. Prince Bull. A Fairy Tale
28. Out of Town
29. Out of the Season
2. SUNDAY UNDER THREE HEADS, AND OTHER TALES, SKETCHES, ARTICLES, ETC.
 1. Sunday under Three Heads—
 1. Dedication
 2. As it is.
 3. As Sabbath Bills would make it
 4. As it might be made
 5. Extraordinary Gazette
 6. Address on the Completion of the First Volume of 'Bentley's Miscellany'
 7. Address on the Completion of the Second Volume of 'Bentley's Miscellany'
 8. Joseph Grimaldi
 9. The Lamplighter Story
 10. International Copyright
 11. The Agricultural Interest
 12. Threatening Letter to Thomas Hood, from an Ancient Gentleman
 13. John Overs
 14. The Early Closing Movement
 15. The Spirit of Chivalry
 16. Crime and Education
 17. Capital Punishment
 18. Address by Charles Dickens to the Cheap Edition of his Works
 19. To be Read at dusk
 20. Address of the English Author to French Public
 21. Dramatic Rights in Fiction
 22. The Earthquake Shock in England
 23. In Memoriam: W. M. Thackeray
 24. Adelaide Anne Protector
 25. History of 'Pickwick'
 26. The Great International Waling-Match
 27. Chauncy Hare Townshend
 28. On Mr Fechter's Acting

35. Miscellaneous Papers: From The Examiner, Household Words and All the Year Round—Plays and Poems. VOL. I

1. INTRODUCTION
2. MISCELLANIES FROM 'THE EXAMINER' 1838-1849
 1. The Restoration of Shakespeare's 'Lear' to the Stage Scott and his Publishers—I
 2. The Restoration of Shakespeare's 'Lear' to the Stage Scott and his Publishers—II

3. International Copyright
 4. Macready as 'Benedick'
 5. Report of The Commissioners appointed to inquire into the Condition of the Persons variously engaged in the University Of Oxford
 6. Ignorance and Crime
 7. The Chinese Junk
 8. Cruikshank's 'The Drunkard's Children'
 9. The Niger Expedition
 10. The Poetry of Science
 11. The American Panorama
 12. Judicial Special Pleading
 13. Edinburgh Apprentice School Association
 14. Leech's 'The Rising Generation'
 15. The Paradise at Tooting
 16. The Tooting Farm
 17. The Verdict for Drouet
 18. 'Virginie' and 'Black-Eyed Susan'
 19. An American in Europe
 20. Court Ceremonies
3. MISCELLANIES FROM 'HOUSEHOLD WORDS' 1850-1859
1. Address in the First Number of 'Household Words'
 2. Announcement in 'Household Words' of the Approaching Publication of 'All The Year Round'
 3. Address in 'Household Words'
 4. The Amusements of the People—I
 5. The Amusements of the People—II
 6. Perfect Felicity
 7. From the Raven in the Happy Family—I
 8. From the Raven in the Happy Family—II
 9. From the Raven in the Happy Family—II
 10. The 'Good' Hippopotamus
 11. Some Account of an Extraordinary Traveller
 12. A Card from Mr. Booley
 13. Mr. Booley's View of the Last Lord Mayor's Show
 14. Pet Prisoners
 15. Old Lamps for New Ones
 16. The Sunday Screw
 17. Lively Turtle
 18. A Crisis in the Affairs of Mr. John Bull
 19. Mr. Bull's Somnambulist
 20. Our Commission
 21. Proposals for a National Jest-Book
 22. A December Vision
 23. The Last Words of the Old Year
 24. Railway Strikes
 25. Red Tape
 26. The Guild of Literature and Art

27. The Finishing Schoolmaster
28. A Few Conventionalities
29. A Narrative of Extraordinary Suffering
30. Whole Hogs
31. Sucking Pigs
32. A Sleep to Startle us
33. Betting-Shops
34. Trading in Death
35. Where we Stopped Growing
36. Proposals for Amusing Posterity
37. Home for Homeless Women
38. The Spirit Business
39. A Haunted House
40. Gone Astray
41. Frauds on the Fairies
42. Things that cannot be Done
43. Fire and Snow
44. On Strike
45. The Late Mr. Justice Talfourd
46. It is not Generally Known
47. Legal and Equitable Jokes
48. To Working Men
49. An Unsettled Neighbourhood
50. Reflections of a Lord Mayor
51. The Lost Arctic Voyagers—I
52. The Lost Arctic Voyagers—II

MISCELLANIES FROM 'HOUSEHOLD WORDS' 1850-1859 (continued)

53. That Other Public
54. Gaslight Fairies
55. Gone to the Dogs
56. Fast and Loose
57. The Thousand and One Humbugs—I
58. The Thousand and One Humbugs—II
59. The Thousand and One Humbugs—III
60. The Toady Tree
61. Cheap Patriotism
62. Smuggled Relations
63. The Great Baby
64. The Worthy Magistrate
65. A Slight Depreciation of the Currency
66. Insularities
67. A Nightly Scene in London
68. The Friend of the Lions
69. Why?
70. Railway Dreaming
71. The Demeanour of Murderer
72. Nobody, Somebody, and Everybody

73. The Murdered Person
74. Murderous Extremes
75. Stores for the First of April
76. The Best Authority
77. Curious Misprint in the 'Edinburgh Review'
78. Well-Authenticated Rappings
79. An Idea of Mine
80. Please to Leave your Umbrella
81. New Year's Day
82. Chips
83. Supposing!

MISCELLANIES FROM 'ALL THE YEAR ROUND' 1859-1869

1. Address which appeared shortly previous to the completion of the Twentieth Volume (1868) of intimating a New Series of 'All The Year Round'
2. The Poor Man and His Beer
3. Five New Points of Criminal Law
4. Leigh Hunt. A Remonstrance
5. The Tattlesnival Bleater
6. The Young Man from the Country
7. An Enlightened Clergyman
8. Rather a Strong Dose
9. The Martyr Medium
10. The Late Mr. Stanfield
11. A Slight Question Op Fact
12. Landor's Life

PLAYS

1. The Strange Gentleman
2. The Village Coquettes
3. Is She His Wife? Or, Something Singular!
4. The Lamplighter
5. Mr. Nightingale's Diary
6. No Thoroughfare

POEMS

1. THE PICKWICK PAPERS (1837)
2. The Ivy Green
3. A Christmas Carol
4. Gabriel Grub's Song
5. Romance (Sam Weller's Song)
6. THE EXAMINER (1841)
7. The fine Old English Gentleman
8. The Quack Doctor's Proclamation
9. Subjects for Painters
10. THE PATRICIAN'S DAUGHTER (1842)
11. Prologue
12. THE KEEPSAKE (1844)
13. A Word in Season

14. THE DAILY NEWS (1846)
15. The British Lion
16. The Hymn of the Wiltshire Labourers
17. LINES ADDRESSED TO MARK LEMON (1849)
18. New Song
19. HOUSEHOLD WORDS (1850-1851)
20. Hiram Power's Greek Slave
21. Aspire
22. THE LIGHTHOUSE (1855)
23. Prologue
24. The Song of the Wreck
25. THE FROZEN DEEP (1856)
26. Prologue
27. THE WRECK OF THE GOLDEN MARY (1856)
28. A Child Hymn
29. ALL THE YEAR ROUND (1859)
30. The Blacksmith

37. Letters and Speeches VOL. I

1. PREFACES
 1. PREFACE TO THE ORIGINAL EDITION
 2. PREFACE TO THE 'THE CHARLES DICKENS' EDITION
 3. PREFACE TO THE 1893 EDITION
2. THE LETTERS OF CHARLES DICKENS 1833-1870
 1. BOOK I.
 2. BOOK II.
 3. BOOK III.

38. Letters and Speeches VOL. II

- THE LETTERS OF CHARLES DICKENS 1833-1870
 BOOK III (continued)
 THE SPEECHES OF CHARLES DICKENS 1841-1870
1. Edinburgh: June 25, 1841
 2. United States: Jan. 1842
 3. Boston: Feb. 1, 1842
 4. Hartford: Feb. 7, 1842
 5. New York: Feb. 18, 1842
 6. Manchester: Oct. 5, 1843
 7. Liverpool: Feb. 26, 1844
 8. Birmingham: Feb. 28, 1844
 9. London: April 6, 1846
 10. Leeds: Dec. 1, 1847
 11. Glasgow: Dec. 28, 1847
 12. London: Feb. 6, 1850
 13. London: March 1, 1851
 14. London: April 14, 1851
 15. London: May 10, 1851

16. London: June 9, 1851
17. London: June 14, 1852
18. Birmingham: Jan. 6, 1853
19. London: April 30, 1853
20. London: May 1, 1853
21. Birmingham: Dec. 30, 1853
22. London: Dec. 30, 1854
23. Drury Lane: June 27, 1855
24. Sheffield: Dec. 22, 1855
25. London: March 12, 1856
26. London: Nov. 5, 1857
27. London: Feb. 9, 1858
28. Edinburgh: March 26, 1858
29. London: March 29, 1858
30. London: April 29, 1858
31. London: May 1, 1858
32. London: May 8, 1858
33. London: July 21, 1858
34. Manchester: Dec. 3, 1858
35. Coventry: Dec. 4, 1858
36. London: March 29, 1862
37. London: May 20, 1862
38. London: April 12, 1864
39. London: May 11, 1864
40. London: May 9, 1865
41. London: May 20, 1865
42. Knebworth: July 29, 1865
43. London: Feb. 14, 1866
44. London: March 28, 1866
45. London: May 7, 1866
46. London: June 5, 1867
47. London: Sept. 17, 1867
48. London: Nov. 2, 1867
49. Boston: April 8, 1868
50. New York: April 18, 1868
51. New York: April 20, 1868
52. Liverpool: April 10, 1869
53. Sydenham: Aug. 30, 1869
54. Birmingham: Sept. 17, 1869
55. Birmingham: Jan. 6, 1870
56. London: March 15, 1870
57. London: April 5, 1870

London: May 2, 1870

Appendix 6.1: The DCC Keywords List

abashed	adapt	afternoon	aloft
abate	address	afterwards	alone
abbey	adelphi	again	aloof
abear	adieu	agen	aloud
abed	adjoin	aggravate	altar
abel	adjuration	aggravation	alter
abet	adjure	aghost	alteration
abeyance	admirable	agility	alternately
abhorrence	admirably	agin	altogether
abject	admiration	agitate	altro
aboard	admire	agitation	always
abode	admirer	agnes	always
abreast	admonitory	agonised	amaze
abroad	ado	agony	amazement
absorb	adornment	agreeable	amen
abstract	advance	agreeably	amend
abstractedly	adversary	ague	amiability
absurd	adverted	ah	amiable
absurdities	advertise	ahem	amicably
absurdity	advocate	ain	amidst
abundance	afar	air	amiss
abundant	afeerd	airily	among
accent	affability	airy	amount
accidental	affable	alacrity	amphitheatre
acclamations	affec	alarm	ample
accompany	affect	alas	amusement
accomplish	affectation	albany	amy
accord	affection	albion	anastatia
accordingly	affectionate	alderman	and
accursed	affectionately	ale	anecdote
accustom	affery	alehouse	anew
ache	affidavit	alfred	angel
acknowledgment	affirmative	alicia	angrily
acknowledgment	afflict	alicumpaine	angry
s	affliction	alight	animate
acquaint	afford	alike	animosity
acquaintance	affrighted	alive	annie
acquiescence	affront	all	annuity
acquirements	afloat	allan	another
acquit	afoot	allonby	answer
act	afore	allotment	anticipations
actuate	aforesaid	allude	antipathy
acuteness	afraid	allusion	antoine
ada	afresh	alms	anvil

anxious	ark	athenæum	bagman
anxiously	arm	atrocious	bagnet
any	armour	attachment	bagstock
anybody	array	attain	bah
anyhow	arrear	attend	bailey
anything	arter	attendant	baize
anythink	arternoon	attention	baleful
anyways	arterwards	attentive	baliol
anywhere	artful	attentively	ballad
anywheres	artfully	attire	ballantyne
apartment	arthur	attired	balustrade
aperture	article	attitude	bandage
apologetically	articled	auctioneer	bandbox
apology	artifice	audacity	bandy
apoplectic	artless	audible	bangham
apoplexy	as	audibly	banish
apothecary	ascend	aught	banishment
apparition	ascent	augment	banister
appear	ascertain	augustus	banker
appearance	ash	aunt	banquet
appellation	ashamed	auspicious	bantam
append	ashley	autograph	banter
appertain	ashore	autumnal	baps
appetite	ashy	avail	baptist
applaud	aside	avarice	bar
applause	ask	avaricious	barbarous
applicant	aslant	avay	barbary
appoint	asleep	avenger	barber
apprehend	aspect	avert	barbox
apprehension	asperity	avignon	bardell
apprentice	aspire	avocations	bare
apprise	assemblage	avow	bargain
approach	assemble	avowal	barge
approbation	assent	awa	bargeman
approve	assert	await	bargemen
approvingly	assertion	awake	barkis
apron	assiduity	awaken	barmecide
apt	assiduously	awakening	barnaby
arabella	assist	away	barnacle
arbour	assistance	awful	barnet
arcadian	associate	awfully	barney
arch	assurance	awning	barnwell
archway	assure	axiom	baron
ard	assuredly	ay	baronet
ardent	astir	aye	barouche
ardently	astonish	babby	barsad
ardour	astonishment	bachelor	bartholomew
ards	astound	backgammon	base
argumentative	asunder	bad	basely
arise	at	bade	baseness
aristocrat	ath	badger	bashful

basin	belong	bintry	boisterously
basket	beloved	bite	bold
bastille	bench	bitherstone	boldheart
batten	bend	bitter	boldly
bawl	benediction	bitterly	boldness
bayham	benefactor	bitterness	boldwig
bazzard	beneficent	bitzer	bole
be	benevolence	black	bolt
beadle	benevolent	blackfriars	bolter
beadles	bengalee	blacksmith	bonnet
beam	benignant	blameless	bonsor
bear	benignantly	blandois	boodle
bearer	benignity	blank	book
bearing	benjamin	blathers	bookcase
beast	benson	blaze	bookseller
beastly	bequeath	bleater	booley
beau	berry	bless	boot
beautiful	bertha	blessing	booth
beauty	beseech	blessington	bore
beauvais	beset	blight	borrioboola
beaver	beside	blimber	bosom
bebelle	besides	blind	boston
beckon	besmeared	blinder	boswell
beckwith	bespeak	blindness	bottle
bed	bessy	blithe	bouclet
bedford	bestir	blockhead	bough
bedlam	bestow	bloodless	boulogne
bedridden	betake	bloodshed	bound
bedstead	bethink	bloodshot	boulderby
bedwin	betimes	bloom	boundless
befall	betray	bloomer	bountiful
before	betrothed	bloss	bow
beforehand	betsey	blossom	bower
beg	betsy	blot	bowl
beggar	better	blow	bowley
begin	betty	blunder	box
begone	betwixt	blunderbuss	boy
begrimed	bevis	blunderstone	boyish
beguile	bewilder	blush	boythorn
behalf	bewitch	blushingly	boz
behind	bidden	bluster	bradley
behindhand	bidding	board	braid
behold	bidly	boarder	brandy
beholden	bids	boast	brass
beholder	bier	boastful	brave
behoof	biler	boat	bravo
being	bilious	boatman	bray
believe	billiard	bob	brazen
bell	billickin	boffin	bread
bella	billsmethi	boil	break
bellow	bintrey	boisterous	breaker

breakfast	buffle	candlestick	cecilia
breast	buffon	candour	ceiling
breath	bull	cane	celestial
breathe	bulldog	canister	cellar
breathless	bullfinch	cannana	cellarman
breech	bulwer	cannot	ensorious
breed	bumble	canon	ceremonious
brentford	bunch	cant	ceremony
brewer	bundle	canterbury	cerjat
bribe	bunsby	canvass	certain
bride	burgundy	cap	certainly
bridegroom	burial	capacious	chadband
bridesmaid	burn	capitally	chafe
bridge	burst	capitol	chain
bridle	bushel	caprice	chair
brig	busily	capricious	chaise
brigand	business	captain	chaises
briggs	bustle	caravan	chamber
bright	but	carboy	chambermaid
brighten	button	carcase	chance
brightly	buxom	care	chancery
brightness	buzfuz	careless	chandler
brilliancy	bye	carelessly	changeless
brim	bygone	carelessness	chap
brimstone	bystander	caress	chapter
bring	cab	careworn	character
brink	cabin	carker	chargeable
brisk	cabman	carpenter	chariot
briskly	cabriolet	carpet	charker
briskness	cadaverous	carriage	charles
brittles	caddy	carrier	charley
broach	caged	carrock	charlotta
broad	calais	carry	charm
broadside	calamity	carstone	charmingly
broadstairs	calcraft	cart	chartist
broadsword	caleb	carton	chase
brogley	calico	carve	chaste
broil	call	carving	chateau
brooch	callings	casby	chatham
brood	calm	casement	chatter
broom	calmly	cask	cheapside
brother	calmness	cast	cheek
brotherhood	calomel	casting	cheer
brotherly	calton	castle	cheerful
brow	calve	cathedral	cheerfully
browdie	camberwell	cattermole	cheerfulness
brownlow	cambric	cauliflower	cheerily
brute	camel	cause	cheerless
bucket	campagna	cavalletto	cheeryble
budden	candidly	cease	cheeseman
buff	candle	ceaseless	cheggs

chemist	citizeness	codlin	compeyson
cherish	civilised	coffin	complacency
chertsey	clap	cogitation	complacent
cherub	clara	coil	complacently
cherubic	clarence	coincidence	complaint
chesney	clarendon	coketown	complexion
chester	clarionet	cold	compliment
chevy	clarkins	coldly	complimentary
chick	clarkson	coldness	comply
chickenstalker	clasp	coliseum	compose
chickweed	clatter	collar	composedly
chiefly	claypole	collation	composure
chigwell	clearness	collector	compound
child	clemency	collegian	comprehend
childers	clench	collegians	compress
childish	clennam	collins	compromise
chillip	cleopatra	colman	compunction
chime	clergyman	colonel	comrade
chimney	clerk	colonnade	conceal
chimneypiece	clerkenwell	comb	concealment
chin	clever	come	conceit
chink	clickits	comely	conceive
chirp	clifton	comer	conciliate
chirrup	cling	comfort	conciliation
chitling	clink	comfortable	conciliatory
chiv	clissold	comfortably	conclude
chivery	cloak	comforter	conclusion
choakumchild	clock	comic	concord
choice	cloisterham	comical	concourse
choke	closet	comin	concur
choleric	cloth	coming	concurrence
chollop	clothe	command	condescend
chop	cloud	commence	condescendingly
christen	clover	commend	condescension
chronometer	clown	commendation	condition
chuckle	cluppins	commendations	condolence
chuckster	clutch	commiseration	conduct
chuff	cly	commodious	conductor
chuffey	coach	common	conference
church	coachman	commoner	confess
churchwarden	coal	commune	confession
churchyard	coarse	communicate	confide
chuzzlewit	coat	communication	confidence
chuzzlewits	coavinses	companion	confidential
cigar	coaxingly	companionship	confidentially
cincinnati	cobb	company	confine
cinder	cobbler	comparatively	confinement
ciphering	cobbs	compass	conflagration
circumlocution	cobweb	compassion	confound
circumstance	cock	compassionate	confuse
circumstantial	cockade	compassionately	confusedly

confusion	contrive	cove	cry
congratulate	convenience	covent	crypt
congratulations	convenient	coverlet	cuff
congregation	conventionalities	covertly	cumbrous
conjurer	conversation	covetous	cunning
conjuror	conversational	crackit	cupid
connect	converse	cradle	cur
connexion	convey	crafty	curdle
conqueror	conveyance	craggs	curiosity
conscience	convict	cram	curious
conscientiously	conviction	cranmer	curl
conscious	convivial	crape	curse
consciousness	convulse	cratchit	cursor
consent	convulsion	cratchits	curtain
consequence	cookery	cravat	curtsey
consequent	coolly	creak	curtseyed
consequently	coolness	creakle	curtseying
conservatory	coom	creature	custom
considerate	copperfield	creditable	cutlet
consideration	copy	credulity	cuttle
consign	cordial	credulous	cymon
consolation	cordiality	creep	dainty
console	cordially	creetur	daisy
conspicuous	corduroy	creeturs	damme
conspiracy	cork	creevy	damp
conspirator	corn	crevice	damsel
conspire	cornelia	cribbage	dance
constable	corner	crimple	dang
constancy	corney	cripple	danger
constantly	cornhill	crisparkle	dangle
consternation	coroner	crockery	dare
constraint	corporal	crook	daresay
contagion	corpulent	cross	dark
contemplate	correspondent	crow	darken
contemplation	corroborate	crowd	darkly
contemplative	cottage	crowl	darkness
contempt	cough	crown	darling
contemptible	could	cruel	darn
contemptuous	counsel	cruelly	darnay
contemptuously	count	cruelty	dart
contend	countenance	crumb	dartle
content	counterfeit	crummies	dash
continually	counterpane	crummeses	dastardly
continuance	countess	crumpet	datchery
contortion	countryman	crumpton	daughter
contract	courage	crumptions	dauphin
contradict	courier	cruncher	davy
contrairy	court	crupp	dawkins
contrary	courtly	crusoe	dawn
contrition	cousin	crust	day
contrivance	coutts	crutch	daybreak

daylight	delight	despondent	disappointment
dead	delightful	desponding	discharge
deadly	delirious	despotism	disclose
deaf	delirium	destitute	discolour
deal	delude	destitution	discomfit
dear	deluge	detain	discomfiture
dearly	delusion	detest	discomposed
deary	demand	detestable	discomposure
debase	demd	detestation	disconcert
debilitate	demeanour	deuce	disconsolate
debtor	demmit	devil	disconsolately
decanter	demnition	devilish	discontented
decay	demon	devonshire	discouragement
decease	demoniacal	devote	discourse
deceased	demonstration	devotion	discover
deceit	dennis	devour	discovery
deceitful	denominated	devoutly	disdainful
deceive	denounce	dexterity	diseased
decidedly	denunciation	dexterous	disengage
decimus	depart	dexterously	disfigure
declaration	departure	diabolical	disgrace
declare	deplere	dick	disguise
decoration	deponent	dickens	dishonesty
dedlock	deportment	dictation	dishonour
dedlocks	depravity	die	disinclination
deed	depreciate	difficulty	disinterested
deem	depreciation	diffidence	disinterestedly
deep	deprive	digger	disinterestedness
deepen	depth	dignify	dislike
deface	deputation	dignity	dismal
defarge	derision	digression	dismally
defaulter	derisive	diligent	dismantle
defective	descend	dim	dismissal
defendant	descried	diminutive	dismount
deference	descry	dimly	disordered
deferentially	describing	dimpled	disparage
defiance	desert	dine	disparagement
deformity	deserter	dingley	dispatch
defresnier	desertion	dingwall	dispel
degenerate	deserve	dingy	disperse
degradation	desire	dinner	dispirited
degrade	desirous	dinnot	dispose
deign	desist	dint	disposition
dejected	desolate	diogenes	dispute
dejection	desolation	direct	disregard
delay	despair	directly	dissensions
deliberation	despairingly	dirty	distance
delicacy	despatch	disagreeable	distant
delicate	despatched	disagreeably	distantly
delicately	despise	disappearance	distempered
delicious	despondency	disappoint	distinctly

distinctness	dover	duet	edson
distinguish	dowager	duke	edward
distinguishable	dowler	dull	edwin
distract	down	dullborough	ee
distress	downcast	dulness	effect
distrust	downfall	dumb	effectual
distrustful	downstairs	dumkins	effectually
distrustfully	doyce	dump	egad
disturb	doze	dungeon	egotistical
diver	dozen	dunstan	eh
diversion	drab	durden	eighteen
divest	drag	durdles	eighteenpence
divine	dragon	durst	eighth
do	dram	dusky	eightpence
doady	drapery	dust	ejaculate
dobbs	drat	dustman	ejaculation
dock	draught	dusty	eke
dockyard	draw	dutiful	elaborately
doctor	drawbridge	duty	elapse
dodger	drawed	dwarf	elbow
dodo	drawer	dwel	elder
dodson	drawing	dwelling	eldest
doen	dread	dwindle	elevate
dog	dreadful	dyer	elicit
doin	dreadfully	eager	eloquence
dolby	dream	eagerly	eloquent
doleful	drearily	eagerness	else
dolefully	dreary	ear	elysées
doll	dress	earl	em
dolly	dressing	earnest	embellish
dolphus	dressmaker	earnestly	embellishment
dombey	drily	earnestness	embrace
dominion	drink	earth	emigrant
don	drivel	earthly	eminence
donkey	droll	earthy	eminent
doodle	drooce	ease	eminently
door	drood	eat	emma
doorway	droop	eatable	emotion
dor	drop	eatables	emphatic
dora	drouet	eatanswill	emphatically
dorrit	drown	eater	empty
dorsetshire	drowneded	eaves	emulation
dost	drowsily	ebenezer	enamoured
dot	drowsy	echo	encamp
dote	drudgery	ecod	enchant
dotheboys	drummle	ecstasy	enchanter
doubledick	drunkard	eddy	enchantress
doubly	drury	edification	encircle
doubt	dubbley	edify	encomiums
doubtful	dubiously	edith	encounter
dounce	dudgeon	edmunds	encouragement

encumber	esther	expedient	fanny
endear	estimable	expedition	far
endearment	estrangement	expend	fareway
endeavour	eugene	expiration	farewell
endow	eulogium	expire	farrier
endurance	eustace	express	farthing
energy	evasion	expression	fascination
enfeebled	evasively	expressive	fast
engage	evening	expressively	fasten
engagement	evenson	expressly	fastening
engender	ever	exquisite	fatal
english	evermore	extinguish	father
englishman	every	extol	fathom
enjoyment	everybody	extraordinarily	fatigue
enlarge	everythink	extraordinary	favour
enlighten	evidence	extravagance	favourable
enough	evidently	extremity	fawn
enraptured	evil	extricate	fear
enrich	evince	exult	fearful
ensign	evolutions	exultation	fearfully
ensue	evremonde	exultingly	feast
enter	exact	eye	feat
entertain	exactness	eyed	feather
entertainer	exalt	eyeglass	fechter
entertainment	examiner	eyelid	feeble
entrap	exasperate	face	feebly
entreat	exceed	facetious	feeder
entreaty	exceedingly	facetiously	feeling
entrust	excellent	facetiousness	feelingly
envelop	except	fact	feenix
envy	excessively	fade	feign
epistle	exchange	fagin	feint
epitaph	excitable	fail	felicity
equable	excite	fain	fell
equanimity	excitement	faint	feller
equipage	exclaim	faintly	fellow
equipages	exclamation	faintness	fellowship
ere	excursion	fair	felon
erect	excuse	fairly	felton
erection	execrations	faithful	fender
err	execute	faithfully	ferocious
errand	execution	fall	ferocity
escape	executions	FALSE	fervent
especial	exertion	falsehood	fervently
esq	exeunt	falstaff	fervour
esquimaux	exhaust	falter	festivity
esquire	exhibit	familiarly	festoon
essay	existence	famished	fetch
establishment	expatiate	fanciful	fetter
esteem	expectation	fancy	fever
estella	expediency	fang	fevered

feverish	flock	forget	frosty
fezziwig	flog	forgive	frown
fidelity	floor	forgiveness	frugal
fie	flopson	fork	fruitful
field	flora	forlorn	fruitless
fiend	florence	formerly	fugitive
fierce	flounceby	formiville	full
fiery	flourish	forsake	fulness
fifer	flower	forster	functionary
fifteenth	floy	forth	fur
fifty	flurried	forthwith	furiously
filer	flush	fortitude	furnish
filial	flute	fortnight	furniture
finching	flutter	fortun	furnival
find	foe	fortunate	furtherance
finding	fog	fortune	fust
fine	fogg	forward	fustian
finery	foggy	foul	gabblewig
fips	folair	foulon	gabelle
fire	fold	foundling	gabriel
firelight	folk	fourpence	gad
fireplace	follower	fourteen	gaffer
fireside	fond	fourteenth	gaiety
firewood	fondle	fourth	gaily
firmness	fool	fourthly	gait
fish	foolish	fowl	gaiter
fisherman	foot	fractious	gal
fist	footboard	fragment	galileo
fit	footman	fragrant	gall
fitfully	footpath	frankness	gallant
fitz	footstep	fraternal	gallantly
fix	footstool	frederick	gallantry
fixedly	footway	freely	gallery
fixem	forasmuch	frequent	galley
fizkin	forbear	frequenters	gallop
flabby	forbearance	fresh	gallows
flam	forbid	freshen	gamekeeper
flamwell	forcible	fresher	gamfield
flanders	forcibly	freshness	gammon
flannel	foreboding	fretful	gamp
flare	forefinger	friar	gaol
flasher	forego	friend	gaoler
flatter	forehead	friendless	garb
flaxen	foreigner	friendly	gargery
fledgeby	foremost	fright	garland
flee	forenoon	frighten	garment
fling	foresee	frightful	garnish
flintwinch	foreshadow	frightfully	garret
flipfield	forewarn	frill	gashford
flit	forfeit	fro	gaskell
flite	forge	frock	gaslight

gasper	giv	gradgrind	grotesque
gate	give	gradually	ground
gateway	glad	grain	grove
gatherer	gladden	grandchild	grovel
gattleton	gladly	grandmamma	growed
gaudy	gladness	grandmarina	growl
gaunt	glance	grandpapa	growlery
gauzy	glare	grasp	grub
gaveston	glass	grate	grudden
gay	glassful	grateful	grudge
gaze	glaze	gratification	grueby
gen	gleam	gratify	gruel
general	glee	grating	gruff
generous	glide	gratitude	gruffly
genial	glimmering	grave	grumble
genie	glimpse	gravely	grummer
genii	glitter	graver	grundy
genius	globe	gravesend	guard
genoa	gloom	gravity	guardian
genoese	gloomily	gravy	guileless
gent	gloomy	grazinglands	guillotine
genteel	glorious	great	guinea
genteelly	glory	greatcoat	gummidge
gentility	gloucester	greatly	gunter
gentle	glove	greatness	guppy
gentlefolks	glow	greedy	gush
gentleman	go	greengrocer	guster
gentlemanly	goad	greenwich	ha
gentleness	gobler	gregsbury	habit
gently	godmother	gretna	habitable
george	godson	grewgious	habitation
georgiana	goggle	gride	habited
georgina	goin	gridiron	habitually
georgy	golden	gridley	hackney
geranium	goldstraw	grief	haggard
get	good	grieve	hah
ghastly	goodchild	grievously	hail
ghost	goodly	griffin	haldimand
ghostly	goodness	grig	half
gianetta	gore	griggins	halfpenny
gibbet	goswell	grim	hallo
giddy	gout	grimaldi	halloa
gigantic	governess	grimwig	ham
gild	governor	grinby	hammer
giles	gowan	grinder	hammersmith
gill	gown	grindstone	hammock
gilt	grace	groan	hand
gimblet	graceful	grog	handel
gin	gracefully	groggles	handford
gird	gracious	grogzwig	handkerchief
girl	graciously	grope	handmaid

handsome	heartiness	hith	hour
handsomely	heartly	hither	house
handsomer	heave	ho	housebreaker
handsomest	heaven	hoard	housekeeper
hang	heaviness	hoarse	housekeeping
hanging	heavy	hoarsely	houseless
hangman	heed	hoary	housemaid
hanover	heedless	hob	hover
hansardadade	heel	hogarth	how
happiness	heep	hoist	howbeit
happy	heer	holborn	howl
hardihood	heerd	hold	hows
hardly	height	holding	howsa
haredale	heightth	holliday	howsoever
hark	heiress	hollo	hubert
harlequin	helena	holloa	hue
harley	helpless	hollow	hugh
harmless	hem	homage	hulk
harmon	henceforth	homely	hum
haroun	her	homeward	humanising
harriet	herbert	homewards	humble
hart	here	hominy	humbly
harthouse	hereabouts	honest	humbug
haste	hereafter	honeythunder	humility
hasten	herein	honour	humm
hastily	heretofore	honourable	humorous
hasty	hereupon	honourably	humour
hat	hermit	hoo	humouredly
hate	hers	hoof	humph
haughtily	herself	hoop	humphrey
haughtiness	hesitate	hooray	hunchback
haughty	hesitation	hooroar	hundred
haunt	hester	hop	hurrah
have	hexam	hope	hurriedly
havin	heyling	hopeful	hurry
havisham	hicks	hopefulness	husband
hawdon	hide	hopeless	hush
hawk	hideous	hornpipe	i
hawkyard	hideously	horrible	iago
hazard	higden	horribly	idle
he	higham	horror	idleness
head	highminster	horse	idler
headstone	hill	horseback	idlest
heap	him	horsehair	if
hear	himself	hortense	ignominious
hearer	hind	horthe	ignominiously
hearing	hinge	hospitable	ignorance
heart	hint	hostess	ignorant
hearted	hippopotamus	hostler	ikey
hearth	his	hôtel	ill
heartily	hissself	hound	illimitable

illustration	imprudent	inexpressible	inseparable
illustrious	impudence	inexpressibly	insinuate
imaginable	impudent	infallible	insolence
imbecile	impulse	infallibly	insolent
imbecility	imputation	infamous	insolvent
imbue	inanimate	infancy	insomuch
imitate	inasmuch	infantine	inspector
imitation	incense	infer	instant
immediately	incessant	inferior	instantly
immense	incessantly	infernal	instinctive
immensely	inclination	infinite	institution
immensity	incline	infinitely	insupportable
immoderately	incoherently	infinitesimal	intellect
immortal	incompatibility	infirmities	intelligence
immovable	inconsiderable	infirmity	intelligible
imp	inconsiderate	inflammation	intent
impair	inconsistency	infliction	intention
impart	inconsistent	informer	intently
impartially	inconstancy	infuse	intercession
impassioned	inconstant	infusion	interchange
impatience	inconvenience	ingenious	intercourse
impatient	inconvenient	ingenuity	interest
impatiently	incorrigible	ingenuous	interfere
impeachment	incumbrance	ingratitude	interment
impel	indebted	inhabitant	interminable
imperatively	indeed	inhuman	intermission
imperceptibly	indefatigable	inimitable	interpose
imperfect	indentures	injudicious	interposition
imperfectly	indescribable	injurious	interrogatively
imperious	indifference	ink	interrupt
impertinence	indifferent	inkstand	interruption
impertinent	indignant	inmate	interval
imperturbable	indignantly	inmost	intimate
impetuosity	indignation	inn	intimation
impetuous	indispensable	innkeeper	into
implacable	indisposition	innocence	intolerable
implore	indisputable	innocent	intoxication
imploringly	indisputably	innocently	intrude
importunate	indistinct	innumerable	intrusion
impossibilities	indolent	inoffensive	intrusted
impossible	indomitable	inquest	invalid
impostor	induce	inquire	invariable
impracticable	inducement	inquiringly	invariably
imprecation	indulge	inquiry	invitation
impress	industrious	inquisition	involuntarily
impressible	ineffectual	inquisitive	involuntary
impression	ineffectually	inscription	inward
imprisonment	inestimable	inscrutable	inwardly
improbable	inexhaustible	insensibility	irksome
improper	inexorable	insensible	iron
impropriety	inexperience	insensibly	ironical

ironmaster	jobling	kerchief	lammle
irradiate	jocose	ket	lammles
irreconcilable	jocosely	ketch	lamp
irrepressible	jocular	kettle	lamplighter
irresistibly	jocularity	key	landlady
irresolute	joe	keyhole	landless
irresolutely	joey	kidderminster	landlord
irresolution	john	kiln	landor
irving	joker	kimber	lane
it	jollity	kimbo	langton
ith	jolly	kimmeens	languidly
itinerant	jolt	kind	languish
ivins	jonas	kindle	lank
ivy	joram	kindly	lantern
jack	jorgan	kindness	lapse
jackman	jorkins	kindred	lar
jaggers	josiah	king	larder
jail	jot	kingsgate	lark
jailer	journey	kinsman	larkins
jarley	jove	kiss	lass
jarndyce	jovial	kit	lassitude
jarsper	joy	kitterbell	last
jasper	joyful	kitty	lastly
jealous	joyfully	klem	lately
jealousy	judas	knag	latter
jeddler	judge	knapsack	laudable
jeer	judiciously	knave	laudanum
jefferson	judy	knavery	laugh
jeffreys	jug	knee	laundress
jellyby	julia	kneel	lausanne
jem	junction	knit	lavinia
jemima	junction	knob	lavish
jemmy	jupe	knock	lavvy
jennings	jury	knocker	lawyer
jenny	juryman	know	lay
jeremiah	jurymen	knowed	layard
jerk	justice	koeldwethout	leaden
jerrold	justly	krook	leamington
jerry	kate	kummauns	leaning
jest	katey	laborious	leap
jew	keen	labour	lear
jewel	keenly	labourer	leastways
jingle	keep	labyrinth	leathern
jiniwin	keeper	lad	leave
jink	keepsake	lady	leaver
jinkins	kenge	ladyship	ledge
jinkinson	kennel	lagnier	ledger
jip	kent	lambeth	ledrook
jist	kenwigs	lame	leech
jo	kenwigses	lament	leer
jobbed	kep	lamentation	leetle

leg	lion	lucifer	malice
legatee	lip	lucetia	malignant
leggings	liquor	luggage	malignity
leghorn	lirriper	lumbey	mama
legibly	listen	lunatic	mamey
legion	listless	lupin	mamie
legislature	lithers	lure	man
leicester	littimer	lustily	manette
leisure	little	lustre	manfully
leisurely	lively	luxuriously	mangy
lemonade	lizzie	ly	mankind
lend	lm	lytton	manly
length	lmen	ma	manner
lengthen	loaf	maclise	mansion
lenville	lobbs	macready	mantalini
ler	lock	macstinger	mantel
lessen	locksmith	mad	mantua
lest	locomotive	madam	maplesone
let	lodge	madame	marchioness
letter	lodger	madden	marguerite
levity	lodging	maddox	marigold
lewsome	lofty	madeira	mariner
liberality	loike	madeline	marion
liberty	loiter	mademoiselle	mark
lie	lonely	madhouse	markleham
lieu	long	madman	marquis
lieutenant	longfellow	madness	marry
light	look	magdalen	marseilles
lightest	lookee	maggy	marsh
lightly	lookers	magistrate	marshalsea
lightness	loom	magnanimous	martha
lightning	loose	magnate	martigny
lightwood	lor	magnificence	martin
lignum	lord	magnificently	marvel
likeness	lordship	magnus	mary
likeways	lorn	magpie	maryon
likewise	lorry	magsman	masey
likings	losberne	magwitch	mass
lilac	lose	mahogany	mast
lilian	loud	maid	master
lillerton	louisa	maiden	masterly
lillyvick	lounge	maitre	mat
limbury	love	majestic	materially
limehouse	loveliness	majestically	matilda
limp	lovetown	majesty	matrimonial
lincoln	lovingly	mak	matrimony
lincolnshire	lowell	make	matron
lineament	lower	maker	matronly
linen	lowness	malady	matter
linger	lowten	malderton	matting
linkinwater	lucie	maldon	may

maylie	merrylegs	missis	motherless
mayn	merrywinkle	mistake	motherly
mayor	messenger	mister	motion
maypole	messieurs	mistress	motionless
me	methodical	mistrust	motive
meagles	metropolis	mistrustful	mould
meagre	micawber	misty	moulder
mean	middlesex	mitton	mouldy
meaner	midnight	mivins	moulineaux
meanest	midshipman	mob	mound
meanly	midst	mockery	mount
meanness	midsummer	moddle	mourn
meantersay	miff	moderation	mourner
meantime	miggs	modestly	mournful
meat	might	modesty	mournfully
mechanically	mightily	moisten	mowcher
meditate	mightn	mollify	mr
meditation	mighty	moloch	mrs
meditative	mild	moment	much
medway	mildewed	momentary	mudfog
meek	mildly	money	muff
meekly	mildness	monflathers	muffin
meekness	mill	monk	muffle
meg	milliner	monmouth	mugby
melancholy	milly	monosyllable	muggleton
mell	milvey	monotonous	mulberry
mellow	mim	monotony	mule
melodiously	mima	monseigneur	mull
melodrama	mincin	monsieur	multitude
meltham	mind	monstrous	mun
memorable	minder	monstrously	munificent
memorandum	mindful	montague	murderer
memorial	mine	montreal	murdstone
menagerie	mingle	monument	murmur
mend	minion	moodily	muse
mender	minister	mooney	musingly
mendicant	ministration	mooninian	musket
mention	minns	moonlight	must
mercantile	minute	moorish	muster
mercenary	mirth	morfin	mustn
merchant	miry	morleena	mute
merciful	miscellany	morn	mutter
mercy	mischance	morning	mutton
merdle	mischief	morose	mutuel
mere	miscreant	morrow	muzzle
merely	miserable	morsel	my
merit	miserably	mortal	myriad
meritorious	misery	mortally	myself
merrily	misfortune	mortification	mysterious
merriment	misgiving	mortimer	mysteriously
merry	miss	mother	mystery

nadgett	nicety	now	oho
nail	nicholas	nubbles	oilskin
name	nickleby	nuisance	old
nancy	nicodemus	number	oliver
nandy	niece	numerous	olliffe
nankeen	niger	nupkins	omer
nap	nigh	nuptial	omit
naples	night	nutmeg	omnibus
nat	nightcap	o	once
nate	nightfall	oaken	oncommon
nathan	nightingale	oakum	one
nathaniel	nightly	oath	onward
natur	nimble	obdurate	ooman
natural	ninepence	obenreizer	ooze
naturally	ninth	obey	open
nature	nipper	obi	opine
natured	nly	object	opinion
naturedly	no	objection	opportune
nay	noah	objectionable	opportunely
near	noakes	obligation	opposite
neat	nobility	oblige	oppress
neatness	noble	obnoxious	oppressive
necessaries	nobleman	obscure	orator
necessity	nobly	obsequiously	oratorical
neck	nobody	observable	organ
neckcloth	nod	observance	originate
neckerchief	noddy	observant	orlick
ned	noggs	observation	ornament
needful	noise	observatory	ornamental
needle	noiseless	observe	orphan
needlework	noiselessly	obstinacy	orsay
needn	noisy	obstinate	orth
neglect	none	obstinately	osgood
negro	nonsense	obtrude	othello
negus	nonsensical	obtrusive	otherest
neighbourhood	nook	occasion	otherwise
neighbouring	noon	occasionally	ottoman
neither	nor	occupation	ought
nell	norah	occupy	oughtn
nelly	norman	occurrence	ourself
nephew	normandy	odd	ouse
nettled	nose	odious	out
neuchatel	nosegay	offal	outcast
never	not	offence	outer
neville	notary	offensive	outlive
new	nothin	officiate	outward
newcome	nothing	oftener	overcast
newgate	nothink	oftenest	overdo
newman	notice	oftentimes	overflow
newsman	notwithstanding	oh	overhang
niagara	novelty	ohio	overhear

overjoyed	parkle	pecksniffs	person
overlook	parley	peculiar	personage
overpower	parlour	peculiarity	perspiration
overseer	paroxysm	peculiarly	persuade
overset	parrot	pecuniary	persuasion
overspread	parson	peecher	perusal
overtone	part	peep	peruse
owe	partake	peepy	peruvian
ower	partiality	peerage	pervade
own	partickler	peerybingle	pervert
oyster	particular	peg	peschiere
pa	parting	peggotty	pestilence
pace	partridge	pegler	pestilent
packer	pass	pelisse	pet
packet	passage	pell	petowker
pail	passenger	pembroke	petticoat
pain	passion	pen	pettifer
painful	passionately	pencil	pettishly
painter	pasteboard	penetrate	pew
pale	pastry	penitence	pewter
palings	pastrycook	penitent	phaeton
pall	pat	penitential	phantom
palpitate	patch	penny	philadelphia
palsied	patent	pensive	philanthropist
pancks	patience	pensively	philanthropy
pane	patiently	pentonville	philosewers
panel	patriarch	people	phœbe
pang	patriot	pep	phraseology
pangloss	patriotic	perceive	phunky
pant	patron	perceptible	physician
pantaloons	patronage	perch	pianoforte
pantomime	patronised	perchance	pickle
pantomimic	patronize	percy	pickleston
pantry	pattens	peremptory	pickpocket
papa	patter	perfect	pickwick
paper	pauper	perfectly	pickwickian
papists	pause	perhaps	pickwickians
parade	pave	peril	picter
paralytic	pavement	periodical	picturesque
parasol	pavilionstone	perish	picturesqueness
parcel	pawkins	perjury	pie
pardiggle	pawnbroker	perker	piece
pardner	payne	perkin	pieman
pardon	peace	permit	pierce
parenthesis	peaceable	perpetual	pig
pariah	peach	perpetually	pigtail
paris	peal	perplex	pillar
parish	peasant	perplexity	pillow
parishioner	peasantry	perseverance	pinch
parisian	pebbleson	persevere	pincushion
parkes	pecksniff	perseveringly	pine

pious	pomp	precipitately	profession
pip	pompous	preface	proffer
pipchin	pon	preliminary	profit
pipe	pond	prematurely	profligate
pipkin	ponder	prentices	profound
pitch	ponderous	preparation	profoundest
pitcher	pony	preparatory	profundity
piteous	pooh	prepare	profuse
piteously	poor	prepossessing	profusion
pitiable	popery	prepossession	promise
pity	popish	preposterous	prompt
placard	poplar	prescription	promptitude
place	porch	present	proof
placid	porrage	presentiment	property
placidly	porochial	presently	propitiate
plaid	portentous	preserve	propitiatory
plain	porter	preserver	propose
plainly	portion	preside	propound
plaintiff	portly	press	proprietor
plant	portmanteau	preston	proprietorship
plate	portrait	presume	propriety
playfulness	possess	pretence	prosper
plaything	possession	pretend	pross
pleasant	possessor	pretension	prostrate
pleasantly	postboy	pretty	prostration
pleasantry	postern	prevail	protector
please	postilion	prevalent	protest
pleasure	postilions	prevent	protestation
plebeian	postman	pride	protracted
plentifully	postmaster	priest	proud
plod	postpone	prig	providence
plorn	postscript	princely	provis
plornish	pot	principal	prowl
pluck	potboy	prison	prudent
plume	potentate	prisoner	prune
plunder	pott	privation	pubsey
plunge	potterson	probable	pudding
ply	pound	proceed	pugstyles
pocket	pour	proceeding	pulpit
podgers	pout	procession	pumblechook
podsnap	powder	proclaim	pump
poetical	praise	procter	punch
pogram	pray	proctor	punctual
point	prayer	proctors	punctually
poising	preach	procure	punishment
poison	precaution	prodgit	pupford
poke	precede	prodigal	pupil
polite	precedent	prodigious	pure
politely	precept	prodigy	purify
politeness	precinct	producibile	purl
polly	precious	profess	purport

purpose	rain	reference	repentance
purposeless	raise	reflection	repentant
purposely	ral	refractory	repetition
purse	rally	refresh	repine
pursue	ralph	refreshment	replenish
pursuer	raly	refuge	reply
pursuit	ramble	regale	repose
puss	ramsgate	regard	repress
put	rank	régnier	reproach
putney	rap	regret	reproachful
putt	rapacious	regularity	reproachfully
pyke	rapidity	rehearsal	reproof
quack	rapture	reign	reprove
quadrangle	rapturously	rejoice	repugnance
quadrille	rascal	rejoicing	repulse
quadruped	rashly	rejoin	repulsive
quail	rattle	rejoinder	request
quaint	ravage	relapse	requisition
quale	raven	relent	resemblance
quality	ravender	relict	reservation
quantity	raving	relief	reserve
quarrel	raybrock	relieve	reside
quart	rayther	relinquish	resign
quarter	razor	relish	resolute
quartern	read	remain	resolution
quebec	readily	remark	resolve
queer	readiness	remarkable	resort
quench	ready	remarkably	resound
querulous	reason	remedy	respect
question	receipt	remember	respectability
questioner	receive	remembrance	respectable
quicken	recess	remiss	respectful
quickness	recklessness	remittance	respectfully
quiet	reclaim	remonstrance	rest
quietly	recoil	remonstrances	restitution
quilt	recollect	remonstrate	restless
quinion	recollection	remorse	restlessly
quit	recommended	remorseful	restlessness
quite	recommend	remote	restoratives
quiver	recompense	remunerative	restore
quotation	reconcile	rend	resume
quoth	recover	render	retainer
r	rectitude	renew	retaliation
rachael	recumbent	renown	retinue
raddle	red	repair	retire
radfoot	redburn	reparation	retort
radiant	redde	repassed	retrace
rag	redeem	repast	retreat
rage	redlaw	repeat	return
railroad	redouble	repel	rev
railway	referable	repent	revel

revenge	rouse	sanguine	secrete
revengeful	rout	sapsea	secretly
reverend	rove	saracen	see
revive	rub	sash	seek
revolve	rubbing	satisfaction	seem
rheumatism	rude	satisfactory	seize
rhine	rudely	satisfy	seldom
ria	rudge	saucepan	selfish
riah	rueful	saucer	selfishly
rich	ruefully	saunter	selfishness
richard	ruffian	savage	seller
rid	rugg	save	semblance
ride	rugged	saviour	sempronius
riderhood	ruin	sawdust	send
ridicule	ruinous	sawyer	sensation
ridiculous	rul	say	sense
rig	rum	saying	sensible
rigaud	ruminare	scadder	sensitiveness
right	rumour	scaffold	sentiment
ring	run	scaley	septimus
ringlet	rush	scanty	sepulchral
rioter	rustic	scarce	seraglio
ripen	rustle	scarcely	serene
rising	rusty	scarli	serenity
roar	ruth	scatter	serjeant
rob	ry	scene	sermon
robber	ryland	scholar	serpent
roche	sacrifice	schoolmaster	servant
rochester	sad	schoolmeaster	serviceable
rock	sagacious	schooner	servile
rockingham	sagacity	scorn	servility
rod	sage	scornfully	ses
rodolph	sail	scotch	settin
roker	sailor	scoundrel	settle
rokesmith	saint	scowl	seventh
rol	sairey	scrap	seventy
roll	saith	scream	sexton
romish	sake	screw	shabbiness
rookery	sala	scrooge	shabby
room	sallow	scruple	shade
rope	sally	scrupulous	shading
rosa	salutation	scull	shadow
rose	salute	scuttle	shadowy
rosebud	sam	sea	shady
rosina	samaritan	seaman	shaggy
rosy	sambo	seasonable	shake
rottenness	same	seat	shall
rouen	samivel	sech	shame
rough	sammy	second	shan
rouncewell	sampson	secrecy	sharp
round	sanguinary	secret	sharpeye

sharpness	simplon	slinkton	so
shave	sincere	slipper	sob
shawl	sincerely	sliverstone	sober
shay	sincerity	sloggins	sofa
shed	sinful	sloop	soft
shelter	sing	slop	soften
sherborne	singe	sloppy	softly
sherry	singly	slouch	soil
shilling	singular	slovenly	sol
shine	singularly	slowboy	soldiery
ship	sink	slowly	solemn
shipwreck	sinner	slumber	solemnity
shoe	sir	slumkey	solemnly
shoemaker	sissy	slurk	solicit
shop	sister	sly	solicitation
shopman	sit	slyly	solicitude
shore	sitch	slyme	solitary
short	sitiwation	slyness	solitude
shortness	sitting	small	solitudes
should	sixpence	smallweed	solomon
shoulder	sixpenny	smangle	sombre
shrewdly	skeleton	smart	some
shriek	sketch	smartness	somebody
shrill	skettles	smauker	somers
shrimp	skewton	smike	sometimes
shrink	skiffins	smile	somewheres
shrivel	skilful	smilingly	son
shut	skilfully	smite	sonorous
shutter	skimpin	smithers	soon
shuttlecock	skimpole	smithfield	sooth
sich	skipper	smock	soothe
sick	skirmish	smooth	soothingly
sickness	skittle	smuggle	sophia
side	skulk	snagsby	sophronia
sidle	skylight	snatch	sophy
sigh	slackbridge	snawley	sordid
sight	slacken	sneer	sorely
signify	slammer	snevellicci	sorrow
signor	slap	snigsworth	sorrowful
signore	slate	snitchey	sorrowfully
sikes	slay	snivel	sorry
silas	slenary	snoady	sort
silence	sleep	snodgrass	soul
silent	sleeper	snore	soundly
silently	sleepy	snort	sour
silk	sleet	snow	southwark
silken	sleeve	snuddin	sou'wester
sim	sliderskew	snuff	sovereign
simile	slight	snuffers	sowerberry
simmun	slily	snug	sownds
simplicity	slink	snuggery	sowship

spacious	stable	stiff	suavity
spar	stag	stiggins	subdue
spare	staggs	stile	subject
spark	stagnant	stilt	submissive
sparkins	stain	stiltstalking	submitively
sparkle	stair	stimulant	subordinate
sparkler	staircase	stipend	subscription
sparrow	stammer	stipendiary	subside
sparsit	stamp	stipulate	subsist
spasmodic	stand	stipulation	subsistence
spatter	stanfield	stir	substitute
speak	stanny	stocking	subterranean
speck	stare	stoker	succeed
speckle	stargazer	stole	such
spectacle	starlight	stomach	sudden
spectator	startle	stone	suddenly
spectre	startop	stony	suddenness
speculate	starvation	stool	suffer
speculations	starve	stoop	suffering
speculator	state	stop	sufficiently
speech	stateliness	stoppage	suffocate
speechless	stately	storehouse	suffuse
speedily	stationary	stormy	suggestive
spell	stationer	stout	sulkily
spenlow	stay	stove	sulky
spike	steadfast	stow	sullen
spiker	steadfastly	straggle	sullenly
spinster	steadfastness	stragglers	sultan
spire	steadily	straightway	sum
spirit	steadiman	strait	summerson
spite	steadiness	strange	summon
splendid	steady	stranger	sunburn
splendour	steal	stratagem	sunday
spoil	stealthily	straudenheim	sundry
spoke	stealthy	straw	sunset
spoon	steam	stray	superannuated
sportive	steamboat	streak	superfluous
spot	steamboats	stream	superintendence
spottletoe	steamer	street	superior
spout	stead	strew	supernaturally
sprightly	steep	strife	superscription
spring	steeple	strike	superstition
sprinkle	steerage	strive	supper
spruggins	steerforth	strong	supplication
spurn	stentorian	strongly	suppose
spy	step	stryver	supposition
spyers	stern	stump	supposititious
squalid	sternly	stunt	suppress
square	sternness	stupefy	sure
squeers	steward	stupendous	surely
squire	stick	stupor	surgeon

surly	tap	thankee	thunder
surmount	tapa	thankful	thurtell
surpass	taper	thankfulness	thus
surprise	tapkins	that	thy
surround	tapley	thaw	tibbs
surtout	tappertit	thay	tickit
survey	tarnish	theatre	tickler
susan	tarpaulin	theatrical	tiddler
suspect	tartar	thee	tide
suspense	tastefully	theer	tidings
suspicion	tattlesnivel	them	tier
swagger	tattycoram	themselves	tiffey
swallow	tavern	then	tigg
swarm	tavistock	thence	tilda
swear	tax	thenceforth	tile
sweedlepipe	tea	theodosius	till
sweep	teapot	there	tilly
sweet	tear	thereabouts	tim
sweetly	teaspoon	therein	timberry
sweetness	teel	thereof	time
swell	telegraph	thereupon	timid
swidger	telescope	therewith	timidly
swiftness	tell	thief	timorous
swill	tellson	thimble	timson
swindler	temeraire	thine	tinder
swiss	temper	thing	ting
switzerland	temperance	think	tink
swiveller	temperate	thirteenth	tinker
swiviller	temple	thirtieth	tinkle
swoon	tempt	thith	tinkler
swosser	tenant	thither	tippins
syllable	tenanted	thoroughfare	tiptoe
symond	tend	thoroughly	tire
sympathising	tender	thot	tis
sympathy	tenderly	thou	titbull
table	tenderness	though	tite
tacker	tenement	thought	titter
tackleton	tennent	thoughtful	tle
tailor	terminate	thoughtfully	to
taint	termination	thoughtfulness	toady
tak	terrace	thoughtless	toast
take	terrific	thousand	tobacco
taker	terror	thowt	toby
takin	testamentary	thquire	toddyhigh
tale	testator	threadbare	todgers
talfourd	testify	threepence	together
talk	testily	thrice	toil
talkative	testimonial	throng	toilette
tallow	testimony	throw	token
tally	tetterby	thrust	tol
tambourine	thank	thumb	tolerable

tolerably	tregarthen	turnkeys	unctuous
tollman	tremble	turnpike	undeceive
tom	tremendous	turret	undecided
tomb	tremulous	turveydrop	undefined
tombstone	trespass	tut	undergo
tomkins	trifle	twelve	underhanded
tone	tripe	twelvemonth	undersigned
tongs	triumph	twemlow	understand
tongue	triumphant	twenty	undertaker
too	triumphantly	twigger	undeserving
toodle	trooper	twilight	undiminished
toot	trophy	twine	undoubted
toothache	trot	twinkle	undress
toothpick	trot	twinkleton	undutiful
tope	trotter	twopence	unearthly
topmost	trotty	twopenny	uneasiness
torch	trotwood	ty	uneasy
tortoise	trouble	tyburn	unfathomable
torture	troublesome	tyrant	unfeeling
toss	TRUE	ud	unfort
totter	truly	ugh	unfortunate
tottle	trumpet	ull	unfrequently
touch	truncheon	umble	ungainly
tow	trundle	umbrella	ungenerous
towards	trust	unabated	ungrateful
tower	trustee	unaccountable	unhappily
towlinson	trustful	unaccountably	unhappiness
town	trustfulness	unaccustomed	unhappy
townsman	trusty	unacquainted	unimpaired
tox	truth	unaffectedly	unimpeachable
tozer	tub	unalterable	uninitiated
trabb	tuckle	unaltered	unintelligible
traddles	tugby	unapproachable	unite
trade	tuggs	unassisted	unjustly
tradesman	tuggses	unavailing	unknown
tradespeople	tulkinghorn	unbecoming	unlimited
traitor	tulrumber	unbeknown	unlooked
tramp	tumble	unbounded	unlucky
trample	tumbler	unbroken	unmeaning
tranquil	tumbrils	uncertain	unmindful
tranquilly	tumult	uncle	unmitigated
transact	tumultuous	uncomfortable	unmixed
transitory	tune	uncommercial	unmoved
travel	tupman	uncommon	unnat
traveller	tupple	uncommonly	unnatural
traverse	turban	uncomplaining	unopened
treachery	tureen	unconditionally	unpleasantness
tread	turk	uncongenial	unpretending
treason	turn	unconscious	unpromising
treasure	turning	unconsciously	unquestionably
treatise	turnkey	uncouth	unquiet

unreason	vagrant	villain	walworth
unreasonable	vague	villainous	wan
unreformed	vain	villainy	wand
unreserved	vainly	vincent	wander
unreservedly	valet	vintner	wanderer
unsay	valise	violent	wandering
unseasonable	valour	violently	wanton
unseen	vane	violoncello	wapping
unsettle	vanish	visage	warbeck
unspeakable	vanity	vision	wardle
unspeakably	vapour	visionary	ware
unsteady	warden	visitation	warehouse
unsuspicious	vault	visitor	warious
untasted	vay	vith	warm
untie	veal	vivacious	warmly
until	veck	vivacity	warn
untiring	vehemence	vixenish	warp
unto	vehement	vizier	washerwoman
unused	vehemently	voice	wasteful
unutterable	veil	voigt	watcher
unvarying	vell	vols	watchful
unwell	veller	volubility	watchfulness
unwholesome	velveteen	volumnia	watchman
unwilling	ven	vould	water
unwillingly	vendale	voyage	waterbrook
unwinking	veneering	voyager	waterloo
unwonted	veneerings	vulgar	waterman
unworthiness	venerable	vulture	watermen
unworthy	eneration	vun	watertoast
up	vengeance	vy	wath
uphill	vent	w	watkins
upholsterer	ventriloquist	wackford	watson
upon	venture	wackles	wave
uppermost	venus	wade	wax
uproar	ver	wafer	way
upshot	vere	wag	waylay
upstairs	verger	wager	wayside
upward	verily	waggon	weak
urbanity	verisopht	waggons	weakness
urge	vermin	wainscot	weal
uriah	very	waistcoat	wear
usage	vessel	waiter	weariness
usher	vestry	wake	wearisome
usual	vex	wakeful	weary
usurer	vexation	wal	weather
utmost	vholes	walk	weathercock
utter	viands	wall	weazen
utterance	vich	wally	wed
vacantly	vigilance	walmers	weedy
vacation	vigour	walter	weel
vagabond	vill	waltz	weep

weevle	whither	windy	wotever
wegg	whittington	wine	would
well	who	winglebury	wouldn
weller	whole	wink	wound
wellington	wholesale	winkle	wozenham
wemmick	wholesome	wintry	wrap
wen	wholly	wipe	wrapper
wench	whom	wisbottle	wrath
werry	whomsoever	wise	wrathful
wery	whoop	wisely	wrayburn
wester	whose	wish	wreath
westlock	whosoever	wistful	wreck
westward	why	wistfully	wren
wet	wi	wit	wrest
wharf	wich	witchem	wretch
what	wicious	with	wretched
whatever	wick	withal	wretchedness
wheel	wickam	withdraw	wring
wheelwright	wicked	wither	wristbands
wheel	wickedness	witherden	write
wheezy	wicket	without	writhe
whelk	wickfield	wititterly	writin
whelp	widder	witness	wrong
when	widdy	wittles	wry
whence	widger	woe	wy
whenever	widow	wold	yah
whensoever	widower	woman	yard
whereat	wife	womanly	yarmouth
wherefore	wig	wonder	yawn
wherein	wild	wonderful	ye
whereof	wilding	wonderfully	yes
whereon	wildly	wonderingly	yet
wheresoever	wildness	wont	yield
whereupon	wilfer	woodcourt	yo
wherewith	wilful	wooden	yoho
whether	wilfully	woodensconce	yonder
whew	wilkie	wopsle	yore
which	wilkins	word	yorick
whiffler	will	work	yorkshireman
while	willet	workbox	you
whimsical	willin	workhouse	young
whip	willingly	workman	your
whirl	willises	workmanship	yourn
whisker	wills	worldly	yours
whiskered	wilt	worm	yourself
whisper	wind	worrit	youth
whist	winder	worshipful	youthful
whistle	windings	worst	zeal
white	windlass	worthless	zealous
whitechapel	windmill	worthy	zenith
whiten	window	wos	

Appendix 6.2: Three-Word Lexical Bundles of the DCC

Total No. of N-Gram Types: 100,946
173,5319

Total No. of N-Gram Tokens:

Minimum Frequency: 5

Minimum Range: 5

i don t
out of the
one of the
don t know
there was a
it was a
as if he
that he was
that it was
that he had
he had been
i am not
i have been
would have been
it would be
part of the
what do you
if he had
i am sure
it is a
it was not
to be a
at the door
in the same
that i have
it is not
there was no
i have no
i should have
that i am
if he were
there is a
if i had
might have been
not to be
in the world
he was a
in the morning

it was the
of the house
as if it
there is no
to have been
as well as
of all the
up and down
who had been
which he had
for a moment
the name of
i do not
and in the
at the same
the door and
two or three
side of the
that it is
a kind of
his head and
out of his
a great deal
of his own
to be the
not at all
to say that
that she was
that there was
to be sure
by this time
when he had
and that he
if it were
he would have
and it was
it may be
the end of

i have not
it had been
was in the
in his hand
up to the
i tell you
when he was
to look at
a state of
a long time
as soon as
do you think
as if they
the top of
that he would
i had been
i think i
to do it
was to be
a man of
him in the
of the old
and i am
in which he
that i should
and that the
and all the
the course of
t know what
the same time
as it was
in the house
i am very
now and then
and that i
in such a
of the room
the best of

i dare say
in a very
the first time
he could not
but it was
to see the
end of the
he was not
in the course
it would have
to do with
in his own
it will be
was going to
the midst of
is to be
it is the
of it and
a pair of
on the other
to me and
that she had
as if the
in the midst
for the first
as to the
to him and
to the door
a good deal
i will not
i am a
that they were
i am afraid
and there was
as he was
the subject of
he had a
a sort of

of the most
but i am
have been a
i have a
she had been
i have seen
had been a
he had not
but it is
that he is
the part of
this is the
he is a
it to be
up in the
to have a
and it is
it in the
seemed to be
to think of
in a state
the house of
on the ground
on the subject
of the same
here and there
a couple of
and with a
one of these
that he might
to make a
of the day
had not been
if they were
if it had
on the part
one of them
i should like
in the way
his hand and
i think it
it must be
on one side
of the world
down to the
this is a
of such a
was not a
of the great

for the purpose
back to the
in which the
it might be
looking at the
ought to be
to do so
him with a
seemed to have
the head of
to speak to
when it was
in the air
in the street
of the two
to see him
they had been
the room and
the house and
and he was
a man who
in the first
he did not
into the room
must have been
he had no
the bottom of
as it is
it has been
the door of
the way of
which i have
would be a
and when i
and when he
in at the
in his mind
is a very
to be in
the whole of
to be done
of the way
a part of
but i have
that he should
the back of
to know that
at that time
but he was

to be so
he might have
it was in
half a dozen
one of his
was a little
and he had
out of it
if they had
to go to
for a long
him on the
that you are
with the same
a matter of
and down the
corner of the
and on the
the honour of
of my own
should like to
a piece of
look at the
at the time
in all the
in the dark
up at the
a little more
one of those
to think that
so much as
the pleasure of
with a smile
of the night
him to the
i am glad
and if you
when they were
that you have
the man who
and when the
in the room
in the streets
the sound of
take care of
top of the
he was in
was a very
of the whole

to say to
it is to
the face of
in spite of
him to be
which he was
with all the
he had never
man with a
out of a
to the last
was in a
it was very
the side of
in the old
as i am
in a moment
to have the
upon the ground
down in the
on the floor
should have been
he was so
in a corner
in one of
what is the
if he could
in the most
three or four
the middle of
a moment s
i am sorry
of the little
at the bottom
not in the
he would be
and the other
him in his
in the face
say that i
on account of
said in a
to make the
with a very
great deal of
there had been
used to be
at all events
could not have

it on the
ought to have
it to the
all the time
by way of
i have done
if she were
in the night
with him and
so far as
and to be
she was a
the other side
down on the
the cause of
the presence of
and at the
had been in
i have heard
that if he
was a great
was on the
come to the
in the evening
man who had
that is to
there was
something
and that she
but there was
in it and
man in the
over and over
that he has
could not be
of which he
state of mind
there was an
for a little
of one of
the sight of
is in the
that in the
this was the
and then he
to his own
was about to
whom he had
and then the

for the moment
they were all
to the old
him and he
in the middle
it with a
up in a
a young man
in and out
in the least
it was all
a few minutes
at this time
time to time
from time to
he began to
in the city
and of the
him in a
in his pocket
on the top
the world and
down into the
would not have
all this time
have been in
is one of
the way to
it was to
there was nothing
came to the
clock in the
have been the
of an hour
of his life
such a thing
the idea of
he had done
it was so
was such a
and a half
and as he
i am quite
that they had
him that he
to be found
he had had
his face and
it in his

with a great
with his own
a very good
had come to
if there were
as long as
by the fire
he was very
them in the
there is nothing
to come and
to take a
his hat and
in order that
is to say
of their own
a quarter of
all the way
for the time
he looked at
think of it
and so forth
but for the
him when he
as they were
he was the
the eyes of
door of the
o clock in
were in the
have the goodness
you may be
as far as
it as a
were going to
and out of
at the top
by any means
is not the
the time of
which it was
and i don
in the little
of the earth
out in the
early in the
and by the
at this moment
been in the

could have been
from the first
he could have
in the afternoon
it is so
it was that
of which the
would not be
and the old
the death of
the shadow of
on the same
to the great
bottom of the
down upon the
into the street
on the back
the habit of
any of the
in her own
that had been
to be seen
his eyes and
in a few
the ground and
and to the
but he had
in which they
quarter of an
to believe that
of the fire
there was not
to make it
of him and
against the wall
the manner of
and over again
every one of
it should be
to be very
all at once
had been so
of those who
on the first
five and twenty
on the road
for him and
in that way
into the house

of the best
to do and
to which the
if he would
an opportunity of
he was going
i wish i
down by the
in a manner
member of the
look at me
the hands of
a man in
and as the
for the last
he must have
out of her
that was the
the goodness to
to him that
an hour or
it was an
of a great
that part of
a young lady
and made a
have been so
him and the
in front of
it with the
of the man
to be made
his head in
out of their
don t see
of the last
he should be
him if he
in the country
side by side
and began to
in course of
in the
neighbourhood
more or less
there were no
up with a
what it was
a good many

not have been
part of his
to him in
but there is
during the whole
it was his
men and women
the centre of
the night and
was the first
on the head
the street and
to him as
to such a
and when they
be in the
have come to
to the end
to one of
and the two
head of the
they would have
to go and
was a good
place in the
at the gate
knowledge of the
on the occasion
on to the
this is not
which is a
and for the
at a loss
but that he
it is my
thought of the
went on to
as one of
face of the
the voice of
which they had
close to the
even in the
that there are
would be to
and had a
began to think
him at the
the memory of

as to be
it was impossible
to keep the
am sorry to
has been a
on the contrary
seem to be
that is the
went to bed
come out of
he had the
he was at
if you like
on his way
as good as
within a few
came into the
he had left
it from the
his mind to
the time and
have had the
upon it and
but they were
in the name
the air and
on the night
should not be
though he had
from which he
have been to
of the law
many a time
the streets and
had had a
there were a
whom i have
to say the
all my life
it up and
the town and
between them and
here it is
he had got
to be and
it is no
the worst of
for his own
he had to

him as a
in all his
the head and
and from the
master of the
a man to
people in the
to the top
were in a
the place where
with a kind
to the ground
and that his
to come out
up into the
a man and
standing in the
an act of
his mind and
as if a
out in a
that when the
more of the
with the rest
a little and
it was too
know that the
will be the
with the old
away in the
to eat and
and there he
where they were
who would have
of having been
they were the
to get out
but a few
from one of
it was he
out for the
in this state
all these things

Appendix 6.3: Key Clusters in the DCC Compared with the BNC

No.	Keyword	Freq.	Texts	Rc. Freq.	Keyness	p
1	josiah bounderby of coketown	23	1	0	130.56	0.0000000000
2	sir edward bulwer lytton	23	2	0	130.56	0.0000000000
3	the bleater's london correspondent	18	1	0	102.17	0.0000000000
4	messrs pyke and pluck	18	1	0	102.17	0.0000000000
5	the strong minded woman	16	1	0	90.82	0.0000000000
6	how d ye do	12	2	0	68.12	0.0000000000
7	ran away a negro	11	1	0	62.44	0.0000000000
8	plashwater weir mill lock	10	1	0	56.76	0.0000000000
9	morning noon and night	15	8	10	52.70	0.0000000000
10	his double eye glass	9	2	0	51.09	0.0000000000
11	sorrow wrong and trouble	9	2	0	51.09	0.0000000000
12	the latin grammar master	9	1	0	51.09	0.0000000000
13	the magic fish bone	9	1	0	51.09	0.0000000000
14	the reverend frank milvey	9	1	0	51.09	0.0000000000
15	for old acquaintance sake	9	8	0	51.09	0.0000000000
16	toby veck toby veck	8	1	0	45.41	0.0000000000
17	tom jack or richard	8	1	0	45.41	0.0000000000
18	a private sitting room	8	4	0	45.41	0.0000000000
19	monsieur	8	1	0	45.41	0.0000000000

	heretofore the marquis					
20	mr la fayette kettle	8	1	0	45.41	0.0000000000
21	the rain fell heavily	7	5	0	39.73	0.0000000000
22	his three cornered hat	7	3	0	39.73	0.0000000000
23	harley street cavendish square	7	1	0	39.73	0.0000000000
24	the u nited states	7	1	0	39.73	0.0000000000
25	my late dear mother	7	2	0	39.73	0.0000000000
26	charles evremonde called darnay	7	1	0	39.73	0.0000000000
27	cornelius brook dingwall esq	7	1	0	39.73	0.0000000000
28	the hard glazed hat	7	1	0	39.73	0.0000000000
29	a shabby genteel man	7	3	0	39.73	0.0000000000
30	two one pound notes	7	1	0	39.73	0.0000000000
31	at drury lane theatre	7	4	0	39.73	0.0000000000
32	tom gradgrind s daughter	7	1	0	39.73	0.0000000000
33	your poor dear papa	7	1	0	39.73	0.0000000000
34	i'll stand by you	7	5	0	39.73	0.0000000000
35	lord decimus tite barnacle	7	1	0	39.73	0.0000000000
36	pounds shillings and pence	11	5	9	36.00	0.0000000001
37	has seen better days	6	5	0	34.06	0.0000000024
38	sir your faithful servant	6	2	0	34.06	0.0000000024
39	railway officers and servants	6	1	0	34.06	0.0000000024
40	hyde park place w	6	1	0	34.06	0.0000000024
41	sir james emerson tennent	6	2	0	34.06	0.0000000024
42	in exact	6	5	0	34.06	0.0000000024

	proportion as					
43	his black velvet cap	6	2	0	34.06	0.0000000024
44	information relative to the	6	5	0	34.06	0.0000000024
45	the inmost recesses of	6	3	0	34.06	0.0000000024
46	sol and captain cuttle	6	1	0	34.06	0.0000000024
47	the associated silent system	6	2	0	34.06	0.0000000024
48	turtle soup and venison	6	2	0	34.06	0.0000000024
49	arthur doyce and clennam	6	1	0	34.06	0.0000000024
50	at arm s length	6	5	0	34.06	0.0000000024
51	my uncle used to	6	2	0	34.06	0.0000000024
52	autumn leaves autumn leaves	6	1	0	34.06	0.0000000024
53	weller said serjeant buzfuz	6	1	0	34.06	0.0000000024
54	your letter this morning	6	2	0	34.06	0.0000000024
55	a confused heap of	6	6	0	34.06	0.0000000024
56	a lone lorn creetur	6	1	0	34.06	0.0000000024
57	with artificial flowers and	6	4	0	34.06	0.0000000024
58	your ladyship says mr	6	1	0	34.06	0.0000000024
59	during this short dialogue	6	4	0	34.06	0.0000000024
60	duke street saint james's	6	1	0	34.06	0.0000000024
61	the rose coloured curtains	6	1	0	34.06	0.0000000024
62	gad s hill place	6	1	0	34.06	0.0000000024
63	quite right said mr	6	6	0	34.06	0.0000000024
64	ever most affectionately yours	6	1	0	34.06	0.0000000024
65	bible and prayer book	6	2	0	34.06	0.0000000024
66	by slow degrees	6	6	0	34.06	0.0000000024

	and					
67	brave lodgings for one	6	2	0	34.06	0.0000000024
68	by very slow degrees	6	6	0	34.06	0.0000000024
69	beheaded on tower hill	6	1	0	34.06	0.0000000024
70	we understand each other	13	7	23	29.47	0.0000000537
71	old red brick house	5	4	0	28.38	0.0000000967
72	presented itself to the	5	5	0	28.38	0.0000000967
73	packet ship the screw	5	1	0	28.38	0.0000000967
74	our mind s eye	5	2	0	28.38	0.0000000967
75	wait till to morrow	5	4	0	28.38	0.0000000967
76	villa dbs moulineaux boulogne	5	1	0	28.38	0.0000000967
77	towers of notre dame	5	3	0	28.38	0.0000000967
78	weather beaten pea coat	5	1	0	28.38	0.0000000967
79	your sister betsey trotwood	5	1	0	28.38	0.0000000967
80	with even more than	5	4	0	28.38	0.0000000967
81	will rally round him	5	2	0	28.38	0.0000000967
82	the domestic economy of	5	5	0	28.38	0.0000000967
83	that miss dorrit is	5	1	0	28.38	0.0000000967
84	sir roger de coverley	5	5	0	28.38	0.0000000967
85	the humble individual who	5	4	0	28.38	0.0000000967
86	the sugar tongs and	5	1	0	28.38	0.0000000967
87	the proof sheets of	5	3	0	28.38	0.0000000967
88	the instrument maker's door	5	1	0	28.38	0.0000000967
89	calm serene and cheerful	5	1	0	28.38	0.0000000967

90	cousin john said ada	5	1	0	28.38	0.0000000967
91	deeply sensible of the	5	5	0	28.38	0.0000000967
92	be two parties to	5	5	0	28.38	0.0000000967
93	boot jack and countenance	5	1	0	28.38	0.0000000967
94	brass plate on the	5	5	0	28.38	0.0000000967
95	for only five minutes	5	3	0	28.38	0.0000000967
96	four miles an hour	5	4	0	28.38	0.0000000967
97	a favourable opportunity of	5	5	0	28.38	0.0000000967
98	do yourself an injustice	5	4	0	28.38	0.0000000967
99	doctor slammer of the	5	1	0	28.38	0.0000000967
100	every soul on board	5	3	0	28.38	0.0000000967
101	a particularly angular man	5	1	0	28.38	0.0000000967
102	ain't o no use	5	4	0	28.38	0.0000000967
103	almost if not quite	5	4	0	28.38	0.0000000967
104	a four post bedstead	5	4	0	28.38	0.0000000967
105	a gentle tap at	5	5	0	28.38	0.0000000967
106	a miscellaneous collection of	5	5	0	28.38	0.0000000967
107	and tim linkinwater's sister	5	1	0	28.38	0.0000000967
108	anniversary festival of the	5	1	0	28.38	0.0000000967
109	articles of wearing apparel	5	4	0	28.38	0.0000000967
110	an angry look at	5	5	0	28.38	0.0000000967
111	an exact account of	5	3	0	28.38	0.0000000967
112	an uneasy sense of	5	4	0	28.38	0.0000000967
113	most honourable privy council	5	4	0	28.38	0.0000000967
114	mr john harmon had	5	1	0	28.38	0.0000000967

115	my high connexion sir	5	1	0	28.38	0.0000000967
116	long legged young man	5	1	0	28.38	0.0000000967
117	low crowned broad brimmed	5	5	0	28.38	0.0000000967
118	more shame for you	5	5	0	28.38	0.0000000967
119	of miss sally brass	5	1	0	28.38	0.0000000967
120	oh indeed said mr	5	4	0	28.38	0.0000000967
121	old fashioned window seat	5	4	0	28.38	0.0000000967
122	never once looked back	5	4	0	28.38	0.0000000967
123	no worse off than	5	4	0	28.38	0.0000000967
124	noble or right honourable	5	1	0	28.38	0.0000000967
125	i protest against it	5	2	0	28.38	0.0000000967
126	in little dorrit's eyes	5	1	0	28.38	0.0000000967
127	in paris when the	5	1	0	28.38	0.0000000967
128	hanover terrace regent's park	5	1	0	28.38	0.0000000967
129	have grown old and	5	4	0	28.38	0.0000000967
130	head was struck off	5	1	0	28.38	0.0000000967
131	know them no more	5	3	0	28.38	0.0000000967
132	lived happy ever afterwards	5	4	0	28.38	0.0000000967
133	long ago when he	5	5	0	28.38	0.0000000967
134	kiss me once more	5	4	0	28.38	0.0000000967
135	it soon appeared that	5	4	0	28.38	0.0000000967
136	keep always at it	5	1	0	28.38	0.0000000967
137	a sufficient reason for	9	7	10	26.01	0.0000003374
138	hanged drawn and quartered	7	3	5	24.04	0.0000009422

Appendix 6.4: The Dickens's Word List (Sample)

a
 a
 an
à
aa
aaa
aaron
aback
abaddeen
abaft
abandon
 abandon
 abandoned
 abandoning
 abandons
 abandonment
abase
 abase
 abased
 abases
 abasing
 abatement
abash
 abashed
abasinthe
abate
 abate
 abated
 abates
 abating
 abatement
abattoir
 abattoirs
abazid
abbaye
abbess
 abbess
 abbesses
abbeville
abbey
 abbey
 abbeys
abbot
 abbot
 abbots

abbotsford
abbreviate
 abbreviate
 abbreviation
 abbreviations
abby
abd
abdicate
abdicate
 abdicated
 abdication
abdomen
 abdomen
abduct
 abduct
 abducted
 abduction
 abductions
abeam
abear
abed
abel
 abels
abelard
aberdare
aberdeen
abernethy
aberration
 aberration
abershaw
aberystwith
abet
 abet
 abets
 abetted
 abetting
 abettings
 abettor
 abettors
abeyance
abhor
 abhor
 abhorred
 abhorring
 abhorrence
 abhorrent
 abhorrently
abide
 abide

abided
abides
abiding
abigail
ability
ability
abilities
abîme
abit
abject
abjectly
abjure
abjure
abjured
abjuration
ablative
ablaze
able
able
abler
ablest
ablution
ablutions
ably
abnegate
abnegating
abnormal
aboard
aboardship
abode
abode
abodes
aboh
aböh
abolish
abolish
abolished
abolishing
abolition
abolitionist
abolitionist
abolitionists
abominate
abominate
abominated
abomination
abomination
abominable
abominably

Appendix 6.5: The Overused Key Concepts in the DCC with Log-Likelihood above 6.63 (99% accuracy)

No.	Item	Key semantic domains	DCCFreq%	BNCSamply%	LL
1	Z8	Pronouns	746664 (13.26)	72023 (7.44) +	25760.47
2	B1	Anatomy and Physiology	70609 (1.25)	5489 (0.57) +	4040.92
3	Q2.1	Speech: Communicative	70083 (1.24)	7024 (0.73) +	2157.25
4	Z99	Unmatched	176400 (3.13)	22165 (2.29) +	2102.64
5	A13.3	Degree: Boosters	51919 (0.92)	4808 (0.50) +	2003.36
6	W2	Light	5467 (0.1)	0 (0.00) +	1735.05
7	M6	Location and Direction	85596 (1.52)	9859 (1.02) +	1578.99
8	S2.2	People: Male	30669 (0.54)	2534 (0.26) +	1550.88
9	A13	Degree	4466 (0.08)	0 (0.00) +	1417.37
10	Z6	Negative	70158 (1.25)	8052 (0.83) +	1314.57
11	E4.1-	Sad	15736 (0.28)	979 (0.10) +	1296.38
12	T1.1	Time: General	4888 (0.09)	47 (0.00) +	1200.67
13	X3.4	Sensory: Sight	30119 (0.53)	2795 (0.29) +	1155.83
14	M1	Moving, Coming and Going	82979 (1.47)	10157 (1.05) +	1142.72
15	X2.2+	Knowledgeable	25924 (0.46)	2302 (0.24) +	1112.44
16	E2+	Like	17633 (0.31)	1372 (0.14) +	1007.36
17	L1+	Alive	4835 (0.09)	93 (0.01) +	968.76
18	E4.2+	Content	7101 (0.13)	352 (0.04) +	768.47
19	W2-	Darkness	2212 (0.04)	0 (0.00) +	702.02
20	M2	Putting, Pulling, Pushing, Transporting	44174 (0.78)	5347 (0.55) +	644.59
21	A5.1+	Evaluation: Good	26710 (0.47)	2905 (0.30) +	621.9
22	X2.6-	Unexpected	3604 (0.06)	100 (0.01) +	607.95
23	X3.2	Sensory: Sound	14091 (0.25)	1271 (0.13) +	581.76
24	E4.1+	Happy	14732 (0.26)	1370 (0.14) +	562.19
25	H2	Parts of Buildings	24034 (0.43)	2642 (0.27) +	537.27
26	G2.2+	Ethical	6493 (0.12)	405 (0.04) +	533
27	S1.2.4+	Polite	3609 (0.06)	130 (0.01) +	515.58
28	X2.1	Thought, Belief	33931 (0.6)	4139 (0.43) +	475.8
29	S7.2+	Respected	5257 (0.09)	316 (0.03) +	453.76
30	T3+	Time: Old; Grown-up	10528 (0.19)	961 (0.10) +	421.74
31	S2.1	People: Female	13629 (0.24)	1364 (0.14) +	421.38
32	O4.2-	Judgement of Appearance: Ugly	8016 (0.14)	660 (0.07) +	408.39
33	N3.2-	Size: Small	12347 (0.22)	1218 (0.13) +	398.9
34	N6+	Frequent	15430 (0.27)	1651 (0.17) +	381.55
35	E3+	Calm	7536 (0.13)	623 (0.06) +	380.63

36	S3.1	Personal Relationship: General	11249 (0.2)	1122 (0.12) +	351.45
37	I1.1-	Money: Lack	4019 (0.07)	242) (0.02) +	346.11
38	O4.2+	Judgement of Appearance: Beautiful	14721 (0.26)	1604 (0.17) +	340.4
39	X3.2-	Sound: Quiet	4695 (0.08)	320 (0.03) +	337.96
40	X2.2-	No Knowledge	4053 (0.07)	252 (0.03) +	334.18
41	T1	Time	16881 (0.3)	1913 (0.20) +	333.89
42	A13.5	Degree: Compromisers	7494 (0.13)	658 (0.07) +	330.41
43	S4	Kin	31173 (0.55)	4002 (0.41) +	326.68
44	X3.2+	Sound: Loud	1839 (0.03)	52) (0.01) +	306.94
45	A8	Seem	12873 (0.23)	1409) (0.15) +	292.59
46	S1.2	Personality Traits	2025 (0.04)	73 (0.01) +	289.13
47	G2.2-	Unethical	6004 (0.11)	516 (0.05) +	278.29
48	N5---	Quantities: Little	1856 (0.03)	65 (0.01) +	270.54
49	B2-	Disease	11656 (0.21)	1275 (0.13) +	265.56
50	M8	Stationary	6664 (0.12)	610 (0.06) +	265.04
51	X3.3	Sensory: Touch	2297 (0.04)	108 (0.01) +	262.17
52	Z4	Discourse Bin	32437 (0.58)	4344 (0.45) +	255.24
53	N5.1+++	Entire; Maximum	830 (0.01)	1 (0.00) +	251.81
54	I1.1+++	Money: Affluence	920 (0.02)	5 (0.00) +	248.99
55	B5	Clothes and Personal Belongings	20879 (0.37)	2625 (0.27) +	248.05
56	X2	Mental Actions and Processes	1500 (0.03)	46 (0.00) +	238.64
57	A13.7	Degree: Minimizers	5096 (0.09)	439 (0.05) +	234.93
58	X1	Psychological Actions, States And Processes	2619 (0.05)	158 (0.02) +	224.97
59	N6+++	Frequent	4423 (0.08)	378 (0.04) +	207.65
60	S7.2-	No Respect	1176 (0.02)	32 (0.00) +	200.52
61	S1.1.4+	Deserving	1404 (0.02)	54 (0.01) +	190.93
62	E1	Emotional Actions, States and Processes General	4265 (0.08)	373 (0.04) +	189.82
63	S1.2.6-	Foolish	2641 (0.05)	184 (0.02) +	183.54
64	X5.2-	Uninterested/Bored/Unenergetic	3492 (0.06)	284 (0.03) +	182.57
65	O4.2	Judgement of Appearance	1280 (0.02)	47 (0.00) +	180.3
66	X2.6+	Expected	6748 (0.12)	714 (0.07) +	173.65
67	A5.2+	Evaluation: True	7214 (0.13)	779 (0.08) +	172.51
68	A1.3-	No Caution	793 (0.01)	13 (0.00) +	168.48
69	N3.8+	Speed: Fast	10610 (0.19)	1282 (0.13) +	156.22
70	A6.2-	Comparing: Unusual	6318 (0.11)	684 (0.07) +	149.65
71	E3-	Violent/Angry	12859 (0.23)	1647 (0.17) +	136.72
72	E5-	Fear/shock	6684 (0.12)	757 (0.08) +	132.53
73	O1.3	Substances and Materials: Gas	3260 (0.06)	300 (0.03) +	127.88
74	X9.2-	Failure	3793 (0.07)	369 (0.04) +	127.69
75	A3+	Existing	151784 (2.7)	24177 (2.50) +	124.69

76	S1.2.3-	Unselfish	980 (0.02)	42 (0.00) +	121.86
77	E6+	Confident	2773 (0.05)	245 (0.03) +	120.45
78	A5.1++	Evaluation: Good	5554 (0.1)	619 (0.06) +	117.61
79	N6-	Infrequent	4024 (0.07)	410 (0.04) +	117.58
80	E5+	Bravery	1821 (0.03)	134 (0.01) +	115.34
81	S3.2	Relationship: Intimacy and Sex	4701 (0.08)	508 (0.05) +	112.12
82	A5.2-	Evaluation: False	3369 (0.06)	332 (0.03) +	109.18
83	X3.4+	Seen	1029 (0.02)	53 (0.01) +	106.91
84	X5.1-	Inattentive	1145 (0.02)	68 (0.01) +	100.41
85	A7-	Unlikely	4347 (0.08)	474 (0.05) +	100.23
86	Q2.2	Speech Acts	62851 (1.12)	9724 (1.00) +	96.57
87	S1.1.3+++	Participating	289 (0.01)	0 (0.00) +	91.72
88	A1.5.1-	Unused	598 (0.01)	20 (0.00) +	89.98
89	S1.2.5-	Weak	1768 (0.03)	146 (0.02) +	89.51
90	I3.2	Work and Employment: Professionalism	890 (0.02)	48 (0.00) +	87.83
91	N5.1+	Entire; Maximum	39230 (0.7)	5944 (0.61) +	85.36
92	E4.1++	Happy	427 (0.01)	9 (0.00) +	82.4
93	T1.1.2	Time: Present; Simultaneous	20500 (0.36)	2961 (0.31) +	82.23
94	X2.5-	Not Understanding	2237 (0.04)	212 (0.02) +	81.11
95	A5.1--	Evaluation: Bad	911 (0.02)	54 (0.01) +	80.09
96	Z7	If	14732 (0.26)	2071 (0.21) +	77.57
97	S1.2.3+	Selfish	2041 (0.04)	191 (0.02) +	76.59
98	G2.1-	Crime	4764 (0.08)	570 (0.06) +	73.67
99	Q3	Language, Speech and Grammar	11930 (0.21)	1653 (0.17) +	71.53
100	A1.1.2	Damaging and Destroying	6348 (0.11)	815 (0.08) +	66.5
101	N3.7+	Long, Tall and Wide	6368 (0.11)	818 (0.08) +	66.49
102	A13.1	Degree: Non-specific	5223 (0.09)	653 (0.07) +	64.07
103	A1.4-	Unlucky	999 (0.02)	73 (0.01) +	64.06
104	N3.8+++	Speed: Fast	432 (0.01)	15 (0.00) +	63.35
105	S3.1-	No Personal Relationship	284 (0.01)	4 (0.00) +	63.33
106	A9	Getting and Giving; Possession	373 (0.01)	11 (0.00) +	60.76
107	X3.1+	Tasty	178 (0)	0 (0.00) +	56.49
108	X7++	Wanted	177 (0)	0 (0.00) +	56.17
109	S1.2.2+	Greedy	1302 (0.02)	117 (0.01) +	54.26
110	F2	Drinks and Alcohol	7137 (0.13)	963 (0.10) +	53.08
111	B2+	Healthy	1879 (0.03)	195 (0.02) +	51.68
112	E2+++	Like	909 (0.02)	74 (0.01) +	47.43
113	Q2.1-	Speech: Not Communicating	537 (0.01)	32 (0.00) +	46.88
114	A3	Being	144 (0)	0 (0.00) +	45.70
115	S5+++	Belonging to a Group	207 (0)	4 (0.00) +	41.4
116	Z1	Personal Names	100544 (1.79)	16434 (1.70) +	36.89
117	G2.1+	Lawful	1169 (0.02)	117 (0.01) +	36.14

118	S1.2.4-	Impolite	1106 (0.02)	109 (0.01) +	35.83
119	O4.6-	Temperature: Cold	2489 (0.04)	303 (0.03) +	35.27
120	S7.1-	No Power	5099 (0.09)	697 (0.07) +	34.24
121	I1.2-	Debt-free	105 (0)	0 (0.00) +	33.32
122	X3.1	Sensory: Taste	1571 (0.03)	176 (0.02) +	32.58
123	A5.1-	Evaluation: Bad	4593 (0.08)	624 (0.06) +	32.39
124	S1.1.4-	Undeserving	101 (0)	0 (0.00) +	32.05
125	N3.5+	Weight: Heavy	1106 (0.02)	113 (0.01) +	32.03
126	S1.2.1+	Informal/Friendly	2286 (0.04)	280 (0.03) +	31.37
127	E6-	Worry	6704 (0.12)	961 (0.10) +	29.14
128	A1.3+	Cautious	2428 (0.04)	305 (0.03) +	28.99
129	O4.6+	Temperature: Hot/On Fire	6819 (0.12)	988 (0.10) +	26.43
130	X5.2+	Interested/Excited/Energetic	10084 (0.18)	1511 (0.16) +	25.79
131	W2--	Darkness	79 (0)	0 (0.00) +	25.07
132	X7.2+	Religion and the Supernatural (S9)	76 (0)	0 (0.00) +	24.12
133	A6.1	Comparing: Similar/Different	401 (0.01)	32 (0.00) +	21.79
134	A5.1--	Evaluation: Bad	870 (0.02)	94 (0.01) +	20.76
135	S3.2-	Relationship: Asexual	117 (0)	3 (0.00) +	20.59
136	A6.1+++	Comparing: Similar	4662 (0.08)	668 (0.07) +	20.36
137	H4-	Non-resident	192 (0)	10 (0.00) +	19.71
138	F3	Smoking and Non-medical Drugs	1147 (0.02)	135 (0.01) +	19.19
139	I1.2+	Spending and Money Loss	59 (0)	0 (0.00) +	18.72
140	X5.1+	Attentive	2524 (0.04)	344 (0.04) +	17.36
141	A15+	Safe	1339 (0.02)	166 (0.02) +	17.22
142	E4.1--	Sad	136 (0)	6 (0.00) +	16.48
143	T1.2	Time: Momentary	8203 (0.15)	1251 (0.13) +	16.13
144	E2-	Dislike	3205 (0.06)	453 (0.05) +	16.03
145	E4.1+++	Happy	183 (0)	11 (0.00) +	15.8
146	Q2.1+	Speech: Talkative	49 (0)	0 (0.00) +	15.55
147	A5.3-	Evaluation: Inaccurate	2494 (0.04)	344 (0.04) +	15.54
148	O4.6+++	Temperature: Hot/On Fire	46 (0)	0 (0.00) +	14.60
149	A13.2	Degree: Maximizers	9307 (0.17)	1439 (0.15) +	14.47
150	S1.1.3	Participation	45 (0)	0 (0.00) +	14.28
151	H5	Furniture and Household Fittings	13072 (0.23)	2062 (0.21) +	13.56
152	N3.7-	Short and Narrow	2231 (0.04)	310 (0.03) +	13.05
153	E3---	Violent/Angry	39 (0)	0 (0.00) +	12.38
154	S1.2.5+	Tough/Strong	3002 (0.05)	433 (0.04) +	12.22
155	A5.2+++	Evaluation: True	38 (0)	0 (0.00) +	12.06
156	A6.2--	Comparing: Unusual	68 (0)	2 (0.00) +	11.09
157	A7++	Likely	54 (0)	1 (0.00) +	10.98
158	N3.2---	Size: Small	216 (0)	18 (0.00) +	10.72

159	W2---	Darkness	32 (0)	0 (0.00) +	10.16
160	N3.3+	Distance: Far	2993 (0.05)	441 (0.05) +	9.54
161	L3-	No Plants	30 (0)	0 (0.00) +	9.52
162	N3.3--	Distance: Near	527 (0.01)	61 (0.01) +	9.51
163	T4	Time: Early/Late	140 (0)	10 (0.00) +	9.34
164	A1.2.4-	Speech Acts (Q2.2)	29 (0)	0 (0.00) +	9.20
165	O4.2+++	Judgement of Appearance: Beautiful	59 (0)	2 (0.00) +	8.8
166	X9.2+++	Success	27 (0)	0 (0.00) +	8.57
167	F2++	Excessive Drinking	445 (0.01)	51 (0.01) +	8.39
168	X3.4-	Unseen	987 (0.02)	131 (0.01) +	8.29
169	A1.9-	Unavoidable	23 (0)	0 (0.00) +	7.30
170	N5.1++	Entire; Maximum	23 (0)	0 (0.00) +	7.30
171	X2.5+	Understanding	3613 (0.06)	551 (0.06) +	7.1
172	A5.1+++	Evaluation: Good	4660 (0.08)	723 (0.07) +	6.8