

The Lexical Syllabus and English Language Coursebooks

A Corpus-Based Case Study of the Highly
Frequent Word *Like*

by
Jason Peppard

A Thesis Submitted to the University of Birmingham for
the Degree of Doctor of Philosophy

Department of English Language and Applied Linguistics
College of Arts and Law
University of Birmingham
January 2021

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis reports on a corpus-based study of the highly frequent word *like*. Taking a lexical syllabus perspective, it aims to answer two main research questions: (1) What are the most frequent uses of *like*, as reflected in a large reference corpus (COCA), and (2) how is *like* treated in English language teaching materials, as reflected in a pedagogic corpus of ELT coursebooks (ELTCC)? In regard to the first research question, prepositional *like* was found to be notably more frequent in the COCA than verbal *like*, with the *BE like* and *VERB like* constructions accounting for the majority of *like* usage. Further analyses of these constructions focus on simile usage and the various functions associated with perception verbs paired with *like*. Next, the problem with traditional lexical and grammatical classification at the single-word level is demonstrated by examining the most frequent intensifying premodifiers, as well as, the comparative and superlative premodifiers used with *like*, and how the traditional approaches to language can often be inadequate, or even misleading due to their oversimplicity, especially when taking phraseology into account. Regarding the second research question, the majority of *like* usage in the ELTCC was found to be verbal rather than prepositional. It is argued that this mismatch between coursebooks and natural language usage of such a highly frequent word can be problematic for language learners, and some recommendations on how to address this discrepancy are discussed.

Dedication

To my loving wife Yuriko, thank you for your endless support and patience. To my two inspirational daughters, Juli and Ami, it's been an amazing journey watching you grow up these last six years. I love you all so much, and I am truly looking forward to all the amazing adventures that we are sure to have together down the road.

Acknowledgements

I would like to take a moment to express my sincere gratitude to the following people, who have helped me in conducting my research, and without whom I would not have been able to complete this thesis. First and foremost, I would like to thank my lead supervisor **Dr Crayton Walker**, for his expert advice and guidance, his detailed and insightful feedback, and of course, for his patience.

I would also like to thank the late **Dave Willis**, for his inspiration and encouragement, and for taking the time to respond to my email queries; **Dr Mark Davies** for developing and maintaining the COCA corpus (among many others) and for personally and quickly responding to all my emails; **Dr Nicholas Groom** (my co-supervisor) and **Dr Jeannette Littlemore** for taking the time to answer my questions and for helping me track down a few hard to find resources; **Dr Bodo Winter** for his help answering my statistics-based questions; **Marc Helgesen** for his valuable feedback and insight into the world of ELT coursebook writing; my mother-in-law, **Atsuko Matsuda** for the countless hours spent looking after the kids; and finally, I would like to thank my good friend **Craig MacDonald**, of クレレ (Currere) Specialist Editorial Support, for his expert editing skills and for saving me countless headaches in dealing with the minutiae of formatting and layout of such a big project. Thank you all for your help!

Contents

Figures	xvii
Tables	xix
Conventions and Abbreviations.....	xxi
Conventions.....	xxi
Abbreviations	xxi
 1: Introduction— <i>Like</i> , Corpus Linguistics, and the Lexical Syllabus.....	1
1.1: The Lexical Syllabus and Corpus Linguistics	1
1.2: The Impact of Corpus Linguistics	3
1.3: <i>Like</i> as the Target Word.....	7
1.3.1: Frequencies of <i>Like</i>	7
1.3.2: Inaccuracies with <i>Like</i>	9
1.3.3: Research Questions	11
1.4: Functions and Senses of <i>Like</i>	12
1.4.1: Similarity and Sameness.....	14
1.4.2: Fondness, Wanting, and Mood	17
1.4.3: Pragmatics.....	18
1.4.4: Summary; <i>Why</i> and <i>What</i>	19
1.5: Organization of This Thesis	19
 2: Framework—Phraseology	21
2.1: Introduction	21
2.2: Phraseological Theories of Language	22
2.3: Two Approaches to Phraseology	27
2.3.1: The Taxonomic Approach	28
2.3.2: The Probabilistic Approach	29
2.3.3: The Approach Taken in This Thesis	29
2.4: Lexicogrammatical Units of Meaning.....	31
2.4.1: Collocation and Colligation.....	32
2.4.2: Phrase	33
2.4.3: Pattern and Framework.....	34
2.5: Identifying Lexical Items.....	37
2.6: Pedagogical Applications	41
2.6.1: The Lexical Syllabus	41

2.6.2: Consciousness-Raising and Data-Driven Learning	44
2.6.3: Implications for Coursebook Syllabus Design	47
2.7: Conclusion	49
 3: Review of Literature—Simile	 50
3.1: Introduction	50
3.2: Defining Simile	51
3.2.1: Simile Components	53
3.2.2: Simile and Literal Comparison	54
3.3: Simile and Metaphor	56
3.3.1: Three Psycholinguistic Models	56
3.3.1.1: The Comparison Model	57
3.3.1.2: The Categorization Model	58
3.3.1.3 The Career of Metaphor Model	60
3.4: The Approach Taken in This Thesis	61
3.5: Types of Simile	61
3.5.1: Classification Schemes	61
3.5.2: Lexicalized and Innovative Similes	64
3.5.3 Analogical Similes	65
3.5.4: Metaphorical Similes	66
3.6: Analyzing Similes	66
3.6.1: Phraseology	67
3.6.2: Recurring Verbs and Semantic Classes	68
3.6.3: Patterns of Target–Vehicle Dissimilarity	70
3.6.4: Functions of <i>Like</i> Similes	71
3.6.5: Simile and Genre	72
3.7 Conclusion	73
 4: Methodology—Corpus-Driven and Corpus-Based	 76
4.1: Introduction	76
4.2: Research Questions	76
4.3: Corpora	78
4.3.1: COCA	78
4.3.2: The ELTCC	83
4.4: Research Methods and Techniques	85
4.4.1: List and Chart Displays	86
4.4.2: Concordance Sampling	88

4.4.3: Collocational Analysis	92
4.4.3.1: Mutual Information for Measuring Collocational Strength	96
4.5: Summary	97
 5: Results—Similes with <i>BE Like</i>	99
5.1: <i>Like</i> as Vehicle for Simile	99
5.2: Frequency of <i>BE Like</i> Similes in the COCA.....	100
5.3: <i>BE Like</i> Simile Type Usage Across Genre	101
5.3.1: <i>BE Like</i> Frequency and Grammatical Structure.....	101
5.3.1.1: Predicative Similes.....	103
5.3.2: Analogous Similes	106
5.3.3: Lexicalized Similes	107
5.3.4: Metaphorical Similes	110
5.4: Target–Vehicle Dissimilarity Across Genres.....	112
5.4.1: Frequency Data for <i>NOUN BE Like NOUN</i>	112
5.4.1.1: Frequency of Simile Usage with <i>NOUN BE Like NOUN</i>	112
5.4.1.2: Most Frequent Target–Vehicle Dissimilarity Patterns	113
5.4.2: Patterns of Target–Vehicle Dissimilarity	115
5.4.2.1: Shared Target–Vehicle Dissimilarity Patterns.....	116
5.4.2.2: Characteristic Target–Vehicle Dissimilarity Patterns.....	121
5.5: Rationale Behind Simile Usage.....	124
5.5.1: Describing the Unfamiliar	124
5.5.2: Creativity and Opinion	125
5.5.3: Elaboration	126
5.6: The Complexities of Simile.....	127
 6: Results—Main Functions of <i>VERB Like</i>	130
6.1: <i>Like</i> with Verbs; Perception.....	130
6.2: <i>PV Like</i> Phraseology.....	133
6.2.1: <i>PV Like</i> Patterning.....	133
6.2.2: <i>PV Like</i> Phrases	136
6.2.3: Shared Collocates	138
6.3: <i>LOOK Like</i>	141
6.3.1: <i>LOOK like</i> Characteristic Collocates	142
6.3.2: <i>LOOK like</i> Usage across Genres	144
6.4: <i>FEEL Like</i>	146
6.4.1: <i>FEEL like</i> Characteristic Collocates	149

6.4.2: <i>FEEL like</i> Usage across Genres	151
6.5: <i>SOUND Like</i>	153
6.5.1: <i>SOUND like</i> Characteristic Collocates	155
6.5.2: <i>SOUND like</i> Usage across Genres	156
6.6: <i>SEEM Like</i>	157
6.6.1 <i>SEEM like</i> Characteristic Collocates	158
6.6.2: <i>SEEM like</i> Usage across Genres	159
6.7: <i>SMELL Like</i>	160
6.7.1 <i>SMELL like</i> Characteristic Collocates	161
6.7.2: <i>SMELL like</i> across Genres	162
6.8: <i>TASTE Like</i>	164
6.8.1: <i>TASTE like</i> Characteristic Collocates	164
6.8.2: <i>TASTE like</i> Usage across Genres	165
6.9: Conclusion	167
 7: Results—Functions of <i>Like</i> When Premodified with <i>Just</i>	 168
7.1: The Problem at the Single-Word Level	168
7.2: Lexical Representation of <i>Just</i> and <i>Like</i>	169
7.2.1 <i>Just</i>	170
7.2.2: <i>Like</i>	171
7.3: Grammatical Classification of Premodifiers	172
7.3.1: Premodifiers Used with <i>Like</i> in the COCA	175
7.3.2: Frequency Distribution of <i>Just Like</i> Functions	178
7.4: <i>Just Like</i> Confirming Prescribed Rules	180
7.4.1: Intensifying <i>Just Like</i>	180
7.4.1.1: <i>Just Like Any/Every/All</i>	180
7.4.1.2: <i>Just Like Always</i>	183
7.4.1.3: <i>People Just Like You</i>	184
7.4.1.4: <i>Just Like You and Me</i>	185
7.4.1.5: <i>Just Like PRON SAY</i>	186
7.4.1.6: <i>Just Like Old Times</i>	187
7.4.1.7: <i>Just Like NOUN (to VERB)</i>	188
7.4.1.8: Sense Verbs and <i>Just Like</i>	189
7.4.2: Focusing <i>Just Like</i>	191
7.5: <i>Just Like</i> Contradicting Prescribed Rules	193
7.5.1: <i>And Just Like That</i>	194
7.5.2: Pragmatic Use of <i>Like</i> ₃	195

7.6: Representation of <i>Just</i> and <i>Like</i> in the CCLD.....	197
7.7: Conclusion	199
 8: Results— <i>Like</i> with Comparative and Superlative Premodifiers.....	200
8.1: <i>Like</i> , Prescription, and Description	200
8.2: Comparative and Superlative Premodifiers	201
8.2.1: Comparative and Superlative Found with <i>Like</i>	202
8.3: Constructions Confirming Prescribed Rules	206
8.3.1: <i>More Like</i>	206
8.3.2: <i>Less Like</i>	208
8.3.3: <i>Most Like</i>	210
8.3.4: <i>Least Like</i>	212
8.3.4.1: <i>(Not) In the Least Like</i>	213
8.3.5: The Repeated Comparative Construction.....	214
8.4: Constructions Without Prescribed Rules	216
8.4.1: <i>As Much Like</i>	216
8.4.2: <i>Enough Like</i>	217
8.4.3: Grammatically Based Lexical Phrases.....	219
8.4.3.1: <i>BE More Like It</i>	219
8.4.3.2: <i>At Least Like</i>	223
8.4.3.3: <i>About As Much Like</i>	224
8.5: Lexicalized Phrases Without Prescribed Rules	224
8.5.1: <i>That/This BE More Like It</i>	224
8.5.2: <i>More or Less Like</i>	228
8.6: The Importance of Phraseology	229
 9: Discussion—Pedagogic Implications and Applications	230
9.1: Comparing the COCA and the ELTCC.....	230
9.2: Corpus Linguistics and ELT Coursebooks.....	231
9.3: Frequency of <i>LIKE</i> in the ELTCC	234
9.4: <i>LIKE</i> Constructions in the ELTCC	237
9.4.1: <i>Like</i> ₂ (<i>Enjoy/Want</i>).....	237
9.4.1.1: <i>Like</i> _{2.1} (<i>Fond Of</i>).....	237
9.4.1.2: <i>Like</i> _{2.2} (<i>Want</i>)	242
9.4.1.3: <i>Like</i> _{2.3} (<i>In the Mood For</i>)	245
9.4.2: <i>Like</i> ₁ (<i>Similarity</i>).....	246
9.4.2.1: <i>Like</i> ₁ (<i>Similar To</i>).....	246

9.4.2.2: <i>Like</i> _{1,3} (<i>For Example</i>)	255
9.4.2.3: <i>Like</i> _{1,2} (<i>Same As</i>)	256
9.4.3: <i>Like</i> ₃ —Pragmatic Use	256
9.4.4 Premodifiers Used with <i>Like</i>	257
9.5: Pedagogic Implications and Applications	260
9.5.1: Implications	260
9.5.2: Applications	262
9.6: The Lexical Syllabus As Solution	265
 10: Conclusion— <i>Like</i> , Corpus Findings, and ELT Coursebooks.....	268
10.1: <i>Like</i> , Corpora, and ELT Coursebooks	268
10.2: Summary of Findings	269
10.2.1: Simile Is Worthy of Further Study	269
10.2.2: Sense Verbs Are Not Always Sense Verbs	273
10.2.3: Traditional Approaches Are Often Inadequate	274
10.2.3.1: <i>Just Like</i>	274
10.2.3.2: <i>Like</i> with Comparative and Superlative Premodifiers.....	276
10.2.4: <i>Like</i> in ELT Coursebooks.....	278
10.3: Strengths and Limitations of This Thesis.....	281
10.3.1: Strengths	281
10.3.2: Limitations	283
10.4: Suggestions for Further Research	285
 Appendix 1	287
Appendix 2	289
References.....	290

Figures

2.1: Granger and Paquot's phraseological spectrum.....	30
2.2: <i>That</i> - clauses DDL exercise	45
4.1: Screenshot of the list display in COCA.....	87
4.2: Screenshot of the COCA chart display.....	88
4.3: Screenshot of the COCA concordance KWIC interface.....	89
4.4: Screenshot of the AntConc concordance (KWIC) interface	90
4.5: Screenshot of concordance results for <i>look like</i> in COCA	93
4.6: Screenshot of AntConc concordance results for <i>like</i>	94
5.1: Overall frequency of similes by section of COCA	101
5.2: Percentage frequency distribution—grammatical <i>BE like</i> similes	102
5.3: Percentage of analogical similes by section of the COCA.....	107
5.4: Percentage of metaphorical similes by section of the COCA.....	111
5.5: Percentage of true simile usage with <i>NOUN BE like NOUN</i>	112
6.1: Frequency of <i>LOOK like</i> by COCA section.....	144
6.2: Functions of <i>LOOK like</i> by COCA section	145
6.3: Frequency of <i>FEEL like</i> by COCA section.....	152
6.4: Functions of <i>FEEL like</i> by COCA section	152
6.5: Frequency of <i>SOUND like</i> by COCA section.....	156
6.6: Functions of <i>SOUND like</i> by COCA section	157
6.7: Frequency of <i>SEEM like</i> by COCA section	160
6.8: Frequency of <i>SMELL like</i> by COCA section.....	163
6.9: Function of <i>SMELL like</i> by COCA section	163
6.11: Frequency of <i>TASTE like</i> by COCA section	166
6.12: Function of <i>TASTE like</i> by COCA section	166

8.1: Percentage distributions—comparative and superlative adverbs	205
8.2: Frequency of comparative <i>more like</i> by corpus section	208
8.3: Frequency of comparative <i>less like</i> by corpus section.....	209
8.4: Frequency of superlative <i>most like</i> in each section of the COCA	211
8.5: Frequency of superlative <i>least like</i> in each section of the COCA.....	213
8.6: Frequency of repeated comparative construction by corpus section	215
8.7: Frequency of <i>as much like</i> , as found in each section of the COCA	216
8.8: Frequency of <i>enough like</i> in each section of the COCA	218
8.9: Frequency of <i>BE more like it</i> in each section of the COCA.....	221
8.10: Frequency of <i>at least like</i> in each section of the COCA.....	223
8.11: Frequency of <i>that/this BE more like it</i> from each section of the COCA....	227
8.12: Frequency of <i>more or less like</i> in the COCA by section.....	229
 9.1: Frequency distributions—three main functions of <i>like</i>	 235
9.2: Percentage of simile usage with <i>BE like construction</i>	248
9.3: Per-million occurrences—six main verbs found in <i>PV like</i> constructions ...	249

Tables

1.1: Summary of entries for <i>like</i> in the OED	13
1.2: Functions and senses of <i>like</i>	14
2.1: Examples of constructions at all levels of language	26
2.2: Sample of terms used to describe recurring sequences of two or more words ...	34
4.1: Genres in the COCA	79
4.2: Frequency data for <i>like</i> and extended sequences of related phrases	81
5.1: Frequency of lexicalized <i>BE like</i> similes	108
5.2: Frequency of most common target–vehicle dissimilarity patterns	114
5.3: Shared and characteristic target–vehicle dissimilarity patterns.	116
5.4: Percentage of <i>BE like</i> similes with target and vehicle in same category	126
6.1: Most frequent verbs occurring with <i>like</i> in COCA.	130
6.2: Frequency of perception verbs in <i>VERB like</i> construction	137
6.3: Occurrences of <i>PV like a million bucks/dollars</i> in COCA	137
6.4: <i>PV like</i> and its shared collocates	139
6.5: Characteristic collocates of <i>LOOK like</i>	143
6.6: Characteristic collocates of <i>FEEL like</i>	150
6.7: Characteristic collocates of <i>SOUND like</i>	155
6.9: Characteristic collocates of <i>SEEM like</i>	159
6.10: Characteristic collocates of <i>SMELL like</i>	161
6.10: Characteristic collocates of <i>TASTE like</i>	165
7.1: Summary of definitions for <i>just</i> in <i>Merriam-Webster</i> dictionary	170
7.2: Summary of definitions for <i>like</i> in M-W	172
7.3: Raw frequency of premodifiers found to occur with <i>like</i>	176
7.4: Frequency distribution of <i>just like</i> across three main functions of <i>like</i>	178
7.5: Frequency of <i>just like any</i> , <i>just like every</i> , and <i>just like all</i> phrases	180
7.6: Frequency of <i>just like PRON SAY</i> phrases in COCA	186
8.1: Comparative and superlative grading options for adjectives	201

8.2: Frequency data—comparative and superlative premodifiers.....	203
9.1: Frequency and per-million distributions of <i>like</i> word forms	234
9.2: Frequency and percentage distributions for functions and senses of <i>like</i> ..	236
9.3: Frequency and per-million distributions for all forms of <i>BE Like</i>	246
9.4: Frequency and per-million distributions—main verbs in <i>PV like</i>	250
9.5: Results of informal classroom research	261

Conventions and Abbreviations

Conventions

CAPITAL LETTERS are used for lemmas—all of the word forms of an item—for example, the lemma *LIKE* represents the word forms *like*, *likes*, *liked*, and *liking*, and also for parts of speech or grammatical functions in lexical construction notation, e.g. NOUN, CLAUSE etc.

Italics are used for words introduced, especially in particular forms (e.g.: The most frequent premodifying intensifier of *like* is *just*).

Examples of corpus data will be referenced in the following way:

(1.6) It's **really like apples** and oranges.
(MAG[Jet])

The number in the first set of parentheses signifies the chapter followed by example number. Next, all samples taken from corpora will be displayed in `courier` font. This is followed by the COCA corpus section abbreviation in parentheses, or corpus abbreviation for non-COCA samples, with the source of the data shown in square brackets. The five COCA sections are: SPOK = spoken, FIC = fiction, MAG = magazine, NEWS = newspaper, and ACAD = academic. Other corpora referenced are BNC = British National Corpus, and CC = English Language Coursebook Corpus.

Abbreviations

Works frequently cited are identified by the following abbreviations, listed in alphabetical order:

BNC = British National Corpus

D&L = Downing and Locke's (2006) *English Grammar*

CCLD = *Collins COBUILD Learner's Dictionary* (2003)

COCA = Corpus of Contemporary American English

ELTCC = English Language Teaching Coursebook Corpus

Hanks = "Similes and Sets" (2005)

Moon = "Simile and Dissimilarity" (2011b)

M-W = Merriam-Webster online dictionary

PV = perception verb

SV = sense verb

Wikberg = "Phrasal Similes in the BNC" (2008)

1

Introduction

Like, Corpus Linguistics, and the Lexical Syllabus

1.1: The Lexical Syllabus and Corpus Linguistics

It has now been over 30 years since Sinclair and Renouf (1988) first proposed the lexical syllabus for language learning. Offering evidence from the then 7.3-million-word Bank of English (BoE) corpus, they argued that the language syllabus should be based on the most frequent words in the language, including their main patterns of usage and most frequent collocations.

While modern corpus linguistics was still in its early stages of development in the late 1980s, it has since grown considerably as a methodology, due in large part to advances in computer technology and big data. These advances have dramatically changed the way we view language, and it is now widely acknowledged that the mental lexicons of native and highly-fluent speakers rely heavily on multi-word phraseological constructions to facilitate real-time language processing (e.g., Sinclair 1991, 2004; Nattinger and DeCarrico 1992; Hoey, 2005; Wray 2000, 2002; Gray and Biber 2015). The growth and influence of corpus linguistics is further evident in that it would be hard to find a recently published dictionary or major reference grammar that did not claim to be corpus-based or corpus-informed (O’Keeffe et al. 2007, xi). However, this widespread acceptance of corpus

linguistics research has not yet been fully incorporated into English language teaching (ELT) coursebooks and materials (Koprowski 2005; Burton 2012; Gavioli 2005; Romer 2006; Peppard 2016). Studies have shown, for example, that it remains common practice for coursebook writers to select grammar and vocabulary items without referring to corpora to identify the most useful items and their most natural patterns of usage (Shortall 2007; Burton 2012; Peppard 2016). Furthermore, the lexical phrases that are included in coursebooks are often selected subjectively and without reference to corpus data (Koprowski 2005).

In this thesis, by examining the highly frequent word *like* from a lexical syllabus perspective, I aim to further explore the perceived mismatch between corpus linguistics research and ELT teaching materials. First, by examining a large reference corpus, the Corpus of Contemporary American English (COCA), I will identify the most frequent functions, patterns of usage, and collocations of *like* to determine how this highly frequent word is used in real, everyday English. Next, I will examine a specialized, pedagogic corpus, the English Language Teaching Coursebook Corpus (ELTCC), that consists of general communicative English coursebooks commonly used in private language schools and universities in Japan, to determine how *like* is treated in ELT teaching materials. By comparing the findings of these two corpora, and providing additional classroom evidence, I will show how failing to apply the principles of the lexical syllabus can be detrimental to language learners and lead to communicative errors. Finally, I will conclude with a discussion on the pedagogic applications of corpus linguistics and make some suggestions on how high-frequency words such as *like* could be better presented in future ELT coursebooks.

1.2: The Impact of Corpus Linguistics

As background to this thesis, the emergence of corpus linguistics has had a major impact on the way linguists collect and analyze data, resulting in “research findings that have much greater generalizability and validity than would otherwise be feasible” (Biber and Reppen 2015, 1). The large amounts of authentic data found in corpora, while providing empirical tools for researchers to check the validity of their intuitions, have enabled researchers to shift away from the earlier practice of relying on intuition and invented data.

Criticism of the reliance on intuition and invented data is found throughout corpus linguistics literature. Sinclair, a pioneer in the field, wrote that “human intuition about language is highly specific, and not at all a good guide to what actually happens when the same people actually use the language” (1991, 4). Trudgill (quoted in Stubbs 1996) acknowledged that significant theoretical progress had been made based on the intuitions of past linguists. However, he also argued that “in the final analysis, if linguistics is not about language as it is actually spoken and written by human beings, then it is about nothing at all” (Trudgill, xi). On the topic of corpus-based and corpus-driven teaching materials, Shortall (2007, 161) noted that “many of the claims made in support of corpus-based approaches are in fact criticisms of materials based on writers’ intuitions.” Moreover, Walker (2011) noted that “the corpus is a very useful reference tool for teachers enabling them to check the many intuitions they may have about the language before passing them on to their learners” (111). “It can also,” he continued, “often be used in the classroom to heighten learners’ awareness of some of the more important lexical features of the language” (Walker, 111).

Numerous analyses have demonstrated the benefits of using corpus data over intuition and invented data. For example, in an early study on verb-intensifier collocations, Greenbaum (1974), “to provide access to the cumulative experience of large numbers of speakers,” attempted to gather data on collocations from different groups of informants but found that people generally disagreed. With an analysis of the words *between* and *through*, Kennedy (1991) showed that native-speaker intuition, including that found in language teaching materials, often fails to adequately explain the subtleties of very frequent structural words. He concluded that “the higher the frequency of words and the more complex their semantic structure, the less reliable our intuitions about their most important functions in text” (Kennedy, 97). Finally, Francis and Sinclair (1994, 197), using ergative verbs as an example, showed that their intuition alone was not capable of determining whether the verb *clarify* is ergative in nature. Only with the aid of corpus data were they able to confirm that *clarify* is indeed ergative.

Possibly the most significant contribution made possible by advances in corpus linguistics is the widespread acceptance of the notion of phraseology and its development as a field of study. While there are various theories and different approaches to phraseology, as well as a “vast and confusing terminology” (Granger and Paquot 2008, 27), the main premise underlying the whole of phraseological research is generally agreed upon. According to this premise, the majority of language use is prefabricated, and consists of multi-word, lexicalized patterns and phrases that are treated as single items in the mental lexicon (e.g., Pawley and Syder 1983; Sinclair 1991, 2004; Nattinger and DeCarrico 1992; Hunston and Francis 2000; Hoey 2005; Wray 2002). Furthermore, corpus-based phraseological research has validated Halliday’s concept of lexicogrammar, which refers to the

interdependent nature of lexis and grammar. This directly counters Chomsky's once dominant transformational-generative grammar which treats grammar and lexis as separate entities with lexis often taking a secondary role in the creation of meaning (O'Keeffe et al. 2007, 60). This corpus-based lexicogrammatical view of language has made it necessary to reconsider what constitutes a unit of meaning, leading researchers to reconsider how children acquire language and how second language learners learn a language (Kennedy 1998, 272).

Another important impact of corpus linguistics research comes from the development and growth of spoken corpora. Carter (2007) noted that we know a lot more about written language than we do about spoken language. While this is the case, the emergence of spoken corpora is helping researchers bridge the knowledge gap between written and spoken language. Such corpora include the CANCODE corpus, the spoken sub-corpora of the British National Corpus, the Bank of English, and the Corpus of Contemporary American English. Corpus studies have uncovered numerous differences between written and spoken English. Studies like these prompted Carter and McCarthy to argue that "language teaching which aims to foster speaking skills and natural spoken interaction should be based upon the grammar of spoken language, and not on grammars which mainly reflect written norms" (2005, 141).

One aspect of language that has been ignored in more traditional grammars based on written English is relational language. O'Keeffe et al. (2017, 159) have defined relational language as language used to support positive relationships (opposed to transactional language used for the exchange of information between interlocutors). With the emergence of spoken corpora, corpus researchers can now highlight and analyze areas of

relational language that have traditionally been bound to pragmatics and conversation analysis, incorporating them into a more comprehensive spoken grammar. These areas include conversational routines, hedging, approximation, discourse marking, and vague language. As features of relational language, many of these have been found among the most frequently occurring multi-word lexical items in spoken corpora (O’Keeffe et al.), but they are relatively uncommon in traditional ELT coursebooks. Unsurprisingly (considering its remarkably high frequency), the target word of this study, *like*, is found in several areas of relational language. For example, the discourse marking phrase, *and I was like* (to introduce reported speech) occurs more than 2,000 times in the COCA. *Like* is also commonly used in phrases used for vague language: *stuff like that* occurs over 3,300 times in the COCA, *and the like* occurs almost 5,000 times, and *things like that* occurs more than 6,000 times.

There is no doubt that corpus linguistics has had a significant impact on the way we view and study language; the influence of corpus-based research can be found in almost every area of linguistics, applied linguistics, and language-related resources. Kennedy (1998, 91), pointing out that nearly every large contemporary dictionary already claimed to be corpus-based, stated over 20 years ago that it was “almost inconceivable that worthwhile and comprehensive lexical descriptions can be undertaken without a corpus.” However, an increasing number of researchers are pointing out that the use of corpora among teachers and learners is increasing at a surprisingly slow pace, with potentially important corpus-derived information failing to influence curriculum design and coursebook writing (Kennedy 1998; Gavioli 2005; Koprowski 2005; O’Keeffe et. al. 2007; Shortall 2007; Burton 2012; Peppard 2016).

1.3: *Like* as the Target Word

There are two chief, interconnected reasons for choosing *like* as the target word for this study: It is highly frequent, and it is often misused or misunderstood by learners of English. The rationale for this thesis arises from this mismatch, and it leads directly to my research questions.

1.3.1: Frequencies of *Like*

Regarding the first of these reasons, *like* is one of the most frequent words in the English language and, according to the principles of the lexical syllabus, should therefore be a high-priority word for language teaching. According to an analysis of the Oxford English Corpus (OEC), which contains more than two billion words, *like* is the 54th most commonly used word in English (Oxford Dictionaries 2015). However, the OEC frequency list does not specify the parts of speech for the words in the list. The COCA (2020), by contrast, provides a detailed frequency list of the top 60,000 words in the corpus including information on parts of speech. In the COCA frequency list, *like* as a preposition is the 73rd most frequent word; and as a verb, *like* is the 154th most frequent word.¹ Considering that many of the top 100 words in both the OEC and COCA frequency lists consist of short function words, it seems all the more important that very high-frequency content words such as *like* should be covered thoroughly in language teaching materials.

As I will show and discuss thoroughly in this thesis, one of the main reasons for the very high frequency of *like* is the fact that it has many different functions and senses. In addition to its main prepositional and verbal functions, *like* is also used as a conjunction, an adverb, an adjective, and a

1. As a conjunction, *like* is the 762nd most frequent word in the COCA; as an adjective, it is 1,346th; as an adverb, it is 496th; as a noun, it is 4,768th.

noun, and its multifunctionality results in a multitude of patterns, phrases, and collocates. Unfortunately, however, the multifunctionality of high-frequency words such as *like* is not always made apparent in word frequency lists. As previously mentioned, the often-cited OEC word frequency list conflates all functions of *like* into one entry, which could prove problematic for language learners who refer to the list for vocabulary building.

Similar problems concerning conflation of functions and oversimplification of word frequency lists are found in many online resources as well. Six of the top ten results from the query, “most frequent English words,” using the Google search engine, consist of webpages that provide very simple frequency lists with words only. Additional information (such as part of speech, most frequent patterns of usage, or most frequent collocates) is not provided.

Two of the top ten webpages do include additional information, but the entries for *like* on both websites are problematic. The sixth result, a website called Go Natural English (2020) lists the 1,000 most frequent words in English (sourced from the COCA website) and provides example sentences. However, for the two entries of *like* provided (prepositional and verbal), both example sentences use the verbal form of *like*; that is, the first entry does not provide an example of prepositional *like*. Meanwhile, the ninth result, from the website for dictionary.com (accessed September 26, 2020), only lists *like* as one of the most frequent adjectives in English while, according to the COCA, adjectival *like* is the 1,346th most frequent word. Clicking on *like* as found on this website does however bring the user to a full dictionary entry for *like* that includes all its parts of speech, with example sentences. Nonetheless, it appears inaccurate (and it would be confusing for language learners) to only list *like* as a frequently used adjective.

Next, the tenth result of this web search provides a link to a YouTube video titled “100 Most Common Words: Beginner Vocabulary” (Linguamariana 2017). This video does not include *like* in any form. Finally, only one of the top ten websites for English word frequency lists includes additional information that is correct. Wiktionary (2019), an online collaborative dictionary, is the seventh Google result and links to an English frequency list based on TV and movie scripts. While it does not include part of speech information with the initial listing of *like*, a hyperlink brings the user to a full dictionary entry that includes definitions and example sentences for all parts of speech, in addition to etymology, pronunciation and conjugation information.

From a lexical syllabus perspective, these online word frequency lists are problematic and, in most cases, would likely contribute to misunderstandings or errors of usage for language learners concerning *like*. Somewhat surprisingly, the problem of frequency lists not including part of speech, patterning, or collocate information also appeared in both the more academic list and the general learner-oriented list found on the New General Service List website (Browne, Culligan, and Phillips 2013, 2017). Neither of the main websites for these word lists provided this information.

1.3.2: Inaccuracies with *Like*

The second reason for choosing *like* for this study, as mentioned above, is that language learners tend to use it incorrectly, and this may be due to its numerous functions, patterns, and collocates, all contributing to its high frequency. *Like* has been found to cause both production and reception errors for language learners, as I have noted in my own teaching context and in reading the work of other researchers. Research on student errors conducted by Wu (2016) found that students often use polysemous words in only one way.

The following examples of student errors with *like* seem to suggest that the student speakers are unaware of the prepositional use of *like* and of the patterns *look like NOUN* and *no NOUN like NOUN*.

- (1.1) *a. He look as if another person.
 b. He looks like another person.
(1.2) *a. I think there is no actor such as him.
 b. I think there is no actor like him.
(Wu, 7)

Similarly, in classroom research I have conducted (see 9.5.1) in three first-year university classes focused on communicative English, most students misunderstood the question, “What is your hometown like?” Rather than describing their hometowns using adjectives or telling me about things that were in their hometowns, more than half of the students in all three classes told me what they liked in their hometowns; see Examples 1.3–1.5:

- (1.3) I **like** beautiful nature in my hometown.
(1.4) I **like** my hometown because my hometown is safety
 and have good nature.
(1.5) I **like** the beautiful river.

These answers suggest that the students did not understand the question, specifically the prepositional use of *like* in the *what BE NOUN like* construction. It seems more likely that the students who answered incorrectly misunderstood the question to be “What do you like about your hometown?” with reference to the verbal form of *like*. Since high-frequency words such as *like* are often highly frequent due to their numerous functions, patterns, and collocates, it is likely that these productive and receptive errors result from students’ lacking awareness of the prepositional function of *like*.

In addition to the evidence suggesting students of English are often unaware of the prepositional use of *like*, both the prepositional and verbal

forms of *like* have shared patterning regarding premodifiers, which can also potentially lead to misunderstanding and errors for learners. For example:²

- (1.6) It's **really like apples** and oranges.
(MAG[Jet])
- (1.7) "Although, truth be told, I have an IBM. I don't **really like apples**," he said.
(NEWS[SanFranChron])
- (1.8) It's **just like playing** video games.
(FIC[Analog])
- (1.9) I **just like playing**.
(SPOK[NPR_ATC])

Further evidence for the error potential of *like* can be found in the well-known humorous saying, "time flies like an arrow, fruit flies like a banana." This has been used by linguists in examples of syntactic ambiguity (e.g., Burck 1965; de Mey 1982; Pinker 1994), word play such as antanaclassis (e.g., Lebovits 2002) and puns (e.g., Sherry 2007). Pinker (1994, 206) noted that an early computer parser at Harvard in the 1960s found five separate trees for "time flies like an arrow:"

- (1) Time proceeds as quickly as an arrow proceeds, (the intended reading).
- (2) Measure the speed of flies in the same way that you measure the speed of an arrow.
- (3) Measure the speed of flies in the same way that an arrow measures the speed of flies.
- (4) Measure the speed of flies that resemble an arrow.
- (5) Flies of a particular kind, time-flies, are fond of an arrow.

1.3.3: Research Questions

Due to both its high frequency and the high rate of misunderstanding and misuse among my students, I determined that *like* was an ideal candidate

2. With the exception of examples from other researchers, all examples are from the COCA. Each file in the corpus has metadata including the year of publication, section, and source. All examples provided here will be followed by the section and source in brackets.

for this study. In order to facilitate the research process, the following two research questions and one sub-research question were formulated:

- (1) What are the most frequent uses of *like*, as reflected in a large reference corpus (COCA)?
- (2) How is *like* treated in English language teaching materials, as reflected in a pedagogic corpus of ELT coursebooks (ELTCC)?
- (2.1) Are there notable differences in the treatment of *like* between natural language use (as reflected in the COCA) and pedagogic materials (as reflected in the ELTCC)?

1.4: Functions and Senses of *Like*

The OED lists six main entries for *like*. These are summarized in Table 1.1.

The information in Table 1.1 is only a brief summary of *like* usage in the OED, as there is too much information to fit into a single table due to the many sub-senses of *like* for each entry. In addition, the OED is very comprehensive in that it provides detailed historical descriptions and examples for its entries, including obsolete and regional variations. This would certainly be overwhelming for the average language learner, and I also found it to be unsuitable for categorization of data in the present study. For this reason, I decided to form my own categorization rubric using data from the COCA in the preliminary phase of this research. Categorizing and sorting an initial sample of 500 lines from the entire 450-million-word COCA resulted in eight different senses of *like* according to three core functions or meaning groups. The results of this categorization are shown in Table 1.2.³

3. Due to the nature of the corpus software, adjectives ending in *-like* (*like_s* in the OED as shown in Table 1.2) were not included in the search results for *like* and were therefore not included in this analysis. See Moon (2011a) for an analysis and discussion of the *-like* suffix.

Table 1.1: Summary of entries for *like* in the OED, ordered by frequency. Regional and obsolete entries have been omitted. Sub-senses have also been omitted due to space limitations.

	Part of speech	Frequency per million	Etymology	Definition
like₁	adj., adv., conj., prep.	100–999	early Scandinavian	(1) similar, resembling, alike (2) probable, likely and related senses
like₂	verb ₁	100–999	Germanic	to take pleasure in or be pleased by something
like₃	noun ₁	1–9.9	LIKE adj.	something of the same kind as that previously mentioned or implied
like₄	verb ₂	1–9.9	LIKE adj.	to compare; to imitate
like₅	noun ₂	0.1–0.99	LIKE verb ₁	(1) a feeling of regard, preference, or affection for something or someone; a predilection (2) in the context of social media: an expression of approval or support made by clicking on a particular icon
like₆	suffix	not provided	LIKE adj.	(1) forming adjectives with the sense “similar to or of the nature of—,” “characteristic of or befitting—” (2) forming adverbs with the sense “in or after the manner of—,” “so as to resemble—” (3) forming adjectives with the sense “resembling, or characteristic of, a person who or thing which is—; having the appearance of being—” (4) forming adverbs with the sense “in a—manner; with the appearance of being—”

The data in Table 1.2 show that the most frequent usage of *like* in the COCA is concerned with conveying similarity (*like₁*). I identified three distinct senses of this function.

Table 1.2: Functions and senses of *like*, determined by sorting and categorizing a 500-concordance-line sample from the 450-million-word Corpus of Contemporary American English (COCA), including raw frequency and percentage of sample

Function/sense	Frequency	Percentage
<i>like</i> ₁ similarity	385	77%
<i>like</i> _{1.1} similar to	276	55.2%
<i>like</i> _{1.2} for example	56	11.2%
<i>like</i> _{1.3} same as	53	10.6%
<i>like</i> ₂ enjoyment	89	17.8%
<i>like</i> _{2.1} fond of/enjoy	60	12%
<i>like</i> _{2.2} want	28	5.6%
<i>like</i> _{2.3} in the mood for (feel like)	1	0.2%
<i>like</i> ₃ pragmatic	26	5.2%
<i>like</i> _{3.1} filler	15	3%
<i>like</i> _{3.2} discourse marker	11	2.2%

1.4.1: Similarity and Sameness

First, the most frequent use of *like* overall, *like*_{1.1}, is used to mean *similar to* and accounts for over half of all usage of *like* in the sample. The most frequent construction used with this sense of *like* in the sample is *BE like*, as shown in Examples 1.10–1.12.

(1.10) I **am like** a brochure, conceivably, in being shallow, solicitous, and full of clichés.
(ACAD[Poetry])

(1.11) It **was like** someone had tossed a road map at my feet. (MAG[RollingStone])

(1.12) I told you it's **not like** that.
(FIC[SouthernRev])

Initially, the usages of the *BE like* construction in Examples 1.10–1.12 may appear different from one another. In 1.10, *like* is used as a preposition to introduce a simile. Example 1.11 is also a simile, but here *like* is a conjunction and can be glossed with *as if*, and Example 1.12 appears in a negative structure, stating that the speaker disagrees with a statement. It can be argued, however, that the basic meaning of *like* in these examples is the same, as Downing and Locke (2006, 551) stated: The primary function of *like* is to convey similarity of features or character.

The next most frequent usage of *like* in the sample is concerned with the *VERB like* construction with *LOOK*, *FEEL*, *SOUND*, and *SEEM* being the most common recurring verbs, shown in Examples 1.13–1.16.

- (1.13) And she was striking, with enormous blue eyes,
which made her **look like** a model from the
Eastern Bloc.
(FIC[Bk:SocialClimbers])
- (1.14) Sometimes I **feel like** I'm too cynical, but it
was universally held opinion, I think.
(SPOK[NBC_Dateline])
- (1.15) He's rehabbing, and it **sounds like** he's making
good progress.
(NEWS[NYTimes])
- (1.16) I mean, it **seems like** there's got to be more to
it than that. (ACAD[ABAJournal])

Here too, the basic meaning of *like* is related to the similarity of features or character regarding the connected verb. Examples 1.17–1.19 show three different phrases with *like* used to convey vagueness. Here too, the basic meaning of *like* is concerned with similarity of features or character.

- (1.17) It's **kind of like** what they are saying about
Social Security.
(NEWS[Fox_Beck])
- (1.18) People picking on him, **things like that**,
bullying up on him, trying to get money from
him.
(NEWS[ABC_20/20])
- (1.19) Cover with plastic wrap and leave candies **and
the like** in bags.
(MAG[GoodHousekeeping])

Like is also used with numerical expressions to signify approximation, as shown in Example 1.20. This is also a form of vague language and, again, the main function of expressing similarity is evident. For an in-depth discussion of this usage of *like*, see D'Arcy (2006).

- (1.20) It came out to **like 180** bucks. (MAG[Esquire])

The next most frequent *like*₁ sense, *like*_{1.2}, is used for giving examples, as shown in corpus Examples 1.21–1.23. A few different patterns of this usage were found in the sample. The expression *for example* is sometimes used either directly before or after a *like NOUN* construction (1.21), explicitly marking it as an example. Another *like*_{1.2} pattern is *like NOUN and NOUN*, which is used to provide two examples (1.22), and *like NOUN* can also be placed in parentheses (1.23).

- (1.21) But a lot of the tourist areas, **for example, like Phuket**, they're perfectly fine.
(SPOK[CNN_Cooper])
- (1.22) A bat house needs to be attached to a structure **like a barn or a bridge**, she says, and not mounted on a pole.
(News[NYTimes])
- (1.23) Typically, races last up to 10 hours, with up to 10 miles of off-trail wilderness travel and plenty of diversions (**like obstacle courses**) and first-aid tasks.
(MAG[Backpacker])

The third and least frequently used *like*₁ sense in the sample, *like*_{1.3}, is used to show that something is the same as (or shares the qualities of) some other entity. When used to mean *same*, *like* forms a prepositional phrase with a noun or nominal phrase, or *like NOUN* is followed by a simple clause and is flexible in its placement. As can be seen in Examples 1.24–1.26, this usage of *like* is usually framed with a comma or commas if the *like* phrase is placed in the middle of a sentence. In Example 1.26, we see that the adverb *just* is premodifying *like* which could be glossed as *exactly like* or *same as*.

- (1.24) **Like millions of Americans**, Governor Romney has filed for an extension to complete his tax returns because he's waiting for other information to come in from other entities, that he doesn't have control of their forms.

(SPOK[Fox_Sunday])

- (1.25) I hope that happens, because, **like us**, Japan faces similar problems--a dire fiscal situation, aging population.

(MAG[Newsweek])

- (1.26) For every veteran doing Operation Oliver, there's probably 100 **just like them** who are unemployed.

(NEWS[CSMonitor])

I found the distinction between *like*_{1.1}, *similar to*, and *like*_{1.3}, *same as*, fuzzy at times, suggesting more of a continuum between similarity and sameness rather than a clearly defined difference in meaning. Nevertheless, the data suggested a distinction worthy of categorizing these as two distinct senses, a distinction that becomes apparent when attempting to gloss with either *same as* or *similar to*.

1.4.2: Fondness, Wanting, and Mood

Differing from *like*₁, *like*₂ is a mostly verbal grouping, much more clearly defined, with core patterns that differentiate the meanings. First, *like*_{2.1}, is used to mean *fond of* or *enjoy*, as shown in Examples 1.27–1.28.

- (1.27) I **like** to get up there and connect with people.

(NEWS[Denver])

- (1.28) I **didn't like** that too much.

(MAG[BoysLife])

Next, *like*_{2.2} is used to show that one wants something and is always preceded with a modal verb, usually *would*. In 1.29, we can gloss *would you like* with *do you want*, and in 1.30 *I'd like* can be glossed with *I want*.

- (1.29) **Would you like** to join me for lunch?

(FIC[Analog])

- (1.30) If you'll indulge me, **I'd like** to tell you about one of those childhood experiences.

(FIC[FantasySciFi])

The third *like*₂ sense, *like*_{2.3}, is much less frequent than *like*_{2.1} and *like*_{2.2} and involves a lexicalized form of *FEEL like*, which takes on the meaning of *in the mood for*. I found one example (1.31) of this usage in the initial data sample of 500 concordance lines.

(1.31) Don't **feel like** going out tonight? Curl up with
a DVD gift pack from Blockbuster--we're giving
one away every day this month!
(MAG[Redbook])

There is some disagreement in the literature on this usage of *FEEL like*. See Section 7.5.2.1 for a more detailed discussion and some evidence for including it as a sense of *like*₂.

1.4.3: Pragmatics

While the first two meaning groups are found in both written and spoken English. The third group, *like*₃, includes two senses of *like* that fall within the realm of spoken pragmatics. Adolphs and Carter (2003, 49) found that *like* is over five times more frequent in spoken English than written English, and it therefore needs to be taken into account that the 450-million-word COCA consisted of only 20 percent spoken English. This is reflected in the low proportion of instances for pragmatic *like*.

First, *like*_{3.1} involves the usage of *like* as a discourse particle, or filler, as shown in Example 1.32.

(1.32) And so I did--I did a little research to see,
like, you know, how is Head Start doing?
(SPOK[Fox_Susteren])

Second, *like*_{3.2} is used as a discourse marker, with reported speech following a *BE like* construction. It is generally used in casual, spoken English, as shown in Example 1.33, which is taken from an interview in a magazine.

(1.33) I'm **like** "Let's go to dinner!" And he'll say
he's really tired.
(MAG[Cosmopolitan])

For a comprehensive sociolinguistic analysis of the pragmatic functions of *like*, see D'Arcy (2007, 2017).

1.4.4: Summary; *Why* and *What*

In the initial sorting and categorization process of the preliminary data outlined above, it became clear that prepositional *like* (*like*₁) is notably more frequent than both verbal *like* (*like*₂) and pragmatic *like* (*like*₃), as reflected in Table 1.2. When all instances of *like*₁ were isolated and sorted to the left of *like*, four main patterns emerged as being the most frequent. The first involves the *BE like* construction; the second, the *VERB like* construction, most frequently formed with the perception verbs *LOOK*, *FEEL*, *SOUND*, and *SEEM*; the third involves the intensifying premodifier *just*; the fourth, the comparative premodifier *more*.

In keeping with the main principles of the lexical syllabus, priority should be given to the most frequent functions, patterns, and collocations of the most frequent words. For further analyses in this study, I have chosen these four most frequent patterns of *like* usage.

1.5: Organization of This Thesis

In the following chapters I will provide the theoretical and methodological background to the study, detailed analyses and reviews of the results, and an in-depth discussion on the pedagogic implications and applications of the research. The chapters will proceed as follows.

In Chapters 2 and 3, I explore the theoretical background for the two linguistic phenomena that form the main points of departure for analyzing the corpus data. Chapter 2 looks at the broad topic of phraseology and

provides an outline for its underlying theories of language and the various ways it is defined in the literature, followed by definitions of key phraseological terms used in this study. Chapter 3 provides an overview of simile (where *like* is a key component)—particularly, how simile is identified and the terminology used to describe it, in literature covering relationships between simile and metaphor, and types, categories and analyses of simile.

Next, in Chapter 4, I will provide a detailed explanation of the methodology and research techniques used in this study, including descriptions of the corpora used and the various tools employed to analyze the data.

The results of this study are presented in Chapters 5 to 8, showing corpus analyses for four of the most frequent elements in *like* usage. Chapter 5 investigates the *BE like* construction used to form similes. Chapter 6 looks at the *VERB like* construction with a specific focus on the dominant role of the polysemous verbs of perception paired with prepositional *like*. Chapters 7 and 8 both aim to demonstrate the problem with traditional lexical and grammatical classification systems at the single-word level. Chapter 7 examines the functions of *like* with intensifying premodifiers. Chapter 8 examines the functions of *like* with comparative and superlative premodifiers.

Next, Chapter 9 examines the treatment of *like* in ELT teaching materials by comparing the main findings from Chapters 5 to 8 and corresponding data in a specialized pedagogic corpus of ELT coursebooks. This chapter also includes discussions on the pedagogic implications and applications of this comparative analysis. Finally, Chapter 10 provides a summary of the findings and includes discussions on the limitations of the study and suggestions for future corpus-based research on high-frequency words and their treatment in pedagogic materials.

2

Framework

Phraseology

2.1: Introduction

Generative models of language, such as Chomsky's (1957) transformational grammar, have traditionally treated lexis and syntax as separate and mutually exclusive entities, with lexis taking a secondary, and often peripheral, role to grammatical rules. In the last three decades, however, a new model of language has become increasingly prominent in Western linguistics, largely driven by advances in corpus linguistics. This view of language, generally referred to as the *phraseological approach*, views the grammar-vocabulary dichotomy as artificial and invalid, and it can be broadly defined as "the study of the structure, meaning and use of word combinations" (Cowie 1994, 3168). While there are various differing theories of language that incorporate a phraseological view, they all tend to be usage-based and agree that the mental lexicons of language users largely consist of prefabricated multi-word units that are treated as single units to facilitate fluent language processing.

In this chapter, exploring the topic of phraseology, I will elaborate the background and framework for the thesis as a whole. I aim to provide a comprehensive overview of the theories of language underlying the topic of phraseology, and the various ways it is defined and studied in the literature.

First, in Section 2.2, I will examine several theories that support a phraseological view of language, and I will contrast these with the more traditional, generative, view of language that treats syntax and lexis as separate entities. Next, in Section 2.3, I will outline two different approaches to researching phraseology and provide examples from the literature to illustrate these two approaches. In Section 2.4, I will examine phraseological units of meaning and, with a discussion of the numerous definitions used by different researchers, I will address the problem of terminology in the literature. This will be followed by a detailed explanation of the concepts and definitions of collocation, lexical phrases, lexical patterning, and functions, as used in this thesis. Next, in Section 2.5, I will outline the various ways of identifying lexical items, and I will outline the methods that will be used in the current study. Finally, in Section 2.6, I will discuss the pedagogic applications of corpus-based phraseological research, and I will look at two complementary classroom applications of corpus-based research in phraseology, with examples of how they can be applied for my research as carried out in this thesis.

2.2: Phraseological Theories of Language

As I have already mentioned, the traditional generative view of language, also referred to as the *slot-and-filler* model (Sinclair 1991, 110), or as the dictionary-and-grammar model (Taylor 2012), treats lexis and grammar as separate entities. Language is seen as a system of grammatical rules forming a structure with slots where words can be inserted. However, a large and growing body of research, much of it corpus-based, now provides evidence supporting a phraseological view of language, one that does not make a distinction between lexis and grammar (e.g., Pawley and Syder 1983; Willis 1990; Sinclair 1991, 2004; Nattinger and DeCarrico 1992;

Hasan 1996; Hunston and Francis 2000; Hoey 2005; Granger and Meunier 2008). The phraseological approach is holistic, using data from natural text, and it places priority on syntagmatic relations (semantic relationships of words on the horizontal axis; Sinclair 1998, xv). Generative theories, by contrast, tend to use contrived language examples and to prioritize paradigmatic relations (words that can be substituted in their respective slots on the vertical axis).

One of the first theories of language to offer an alternative approach to the dominant generative grammar of Chomsky, systemic functional linguistics (SFL), was developed by Michael Halliday based on Firth's notion of system. In the view of SFL, language is a social semiotic system (i.e., a social system used to make meaning), and this view differs from generative models of language in that it is primarily concerned with language as a social construct.

This emphasis on the social aspect of language is definitive for SFL, as opposed to the idea that language is an innate human faculty; rather, SFL theory states that language is best understood in terms of *meaning choices in context* (Halliday 1976, 1978; Halliday and Matthiessen 2004). Halliday rejected the grammar-vocabulary dichotomy and argued that lexis and grammar are better understood as a single system to convey meaning, acknowledging that “a great deal of discourse is more or less routinised” (1978, 4). In 1961, he coined the commonly used phrase *lexicogrammar*, writing that “the grammarian's dream is . . . to turn the whole of linguistic form into grammar, hoping to show that lexis can be defined as the ‘most delicate grammar’” (quoted in Hasan 1996, 73). Neal (2006) notes that the system networks which form the “heart” of SFL theory “provide an appropriate way to integrate both ‘grammar’ and ‘lexis’ in a unified component: the

‘lexicogrammar’” (143). While SFL places emphasis on paradigmatic relations, it also includes a syntagmatic element that recognizes multi-word units as single items.

The first model of language to emerge from corpus-based research was developed by John Sinclair, a pioneer of contemporary corpus linguistics. In his work with the COBUILD project, a joint venture between the University of Birmingham and Collins publishers, Sinclair (1991) proposed the *idiom principle*. He argued that “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (110). However, Sinclair acknowledged that the idiom principle alone could not account for the whole of language production, and he proposed a second system to complement the idiom principle, which he termed the *open-choice principle*.

This is a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word or a phrase or a clause), a large range of choice opens up and the only restraint is grammaticalness. (109)

Sinclair argued that the majority of text is composed using the idiom principle, with only occasional reliance on the open-choice principle. He wrote:

It thus appears that a model of language which divides grammar and lexis, and which uses the grammar to provide a string of lexical choice points, is a secondary model. It cannot be relinquished, because a text still has many switch points where the open-choice model will come into play. It has an abstract relevance, in the sense that much of the text shows a potential for being analysed as the result of open choices, but the other principle, the idiom principle, dominates. Open-choice analysis could be imagined as an analytical process which goes on in principle all the time, but whose results are only intermittently called for. (113)

More recently, Hoey (2005) proposed the theory of lexical priming, a usage-based linguistic theory that aims to explain the psycholinguistic process underlying the concepts of collocation and colligation (see Section 2.4.1). By providing statistical evidence from a corpus of newspaper English, Hoey showed that language users come to know a word's phraseology through repeated encounters with it in text and discourse. This knowledge of the word's phraseology, and of its usage with regard to genre, style and social context, operates on a subconscious level. This model reverses the traditional dictionary-and-grammar approach, as it argues that lexis is the main building block of language and that grammar is the result of repeated usage.

The last theory of language that I will discuss here, construction grammar (commonly abbreviated as CxG), is a usage-based theory of language within the field of cognitive linguistics. It was originally developed by linguists such as Lakoff (1987) and Fillmore, Kay, and O'Connor (1988) for the purpose of analyzing idioms (Croft 2007, 15). Similar to Sinclair's (1991, 7) notion that there is no distinction between form and meaning, CxG rejects the grammar-lexis dichotomy, instead arguing that all language is created with *constructions*: "a matching of form and meaning at all levels of generalization" (Hunston and Su 2017, 3). In explaining this core view of CxG, Hilpert (2014) stated that

the line between the mental lexicon, containing knowledge of words, and the mental grammar, containing knowledge of rules, becomes increasingly blurry; so much so that Construction Grammarians propose to abandon it altogether. Instead, knowledge of language is seen as a large inventory of constructions, a construct-i-con. (22).

According to Goldberg (1995), one of the main proponents of CxG, the primary criterion of a construction, is that its associated meaning cannot be derived from its components, and this applies to form-meaning pairing at

all levels of language. Goldberg (2006, 5) also notes, however, that fully predictable, compositional items can be stored as constructions “as long as they occur with sufficient frequency”. Table 2.1 below, reproduced from Goldberg (2006, 5) exemplifies the full spectrum of constructions with examples from all levels of language, from morpheme to clause.

Table 2.1: Examples of constructions at all levels of language. Reprinted from Goldberg (2006, 5)

Morpheme	e.g. <i>pre-</i> , <i>-ing</i>
Word	e.g. <i>avocado</i> , <i>anaconda</i> , <i>and</i>
Complex word	e.g. <i>daredevil</i> , <i>shoo-in</i>
Complex word (partially filled)	e.g. [N-s] (for regular plurals)
Idiom (filled)	e.g. <i>going great guns</i> , <i>give the Devil his due</i>
Idiom (partially filled)	e.g. <i>jog <someone's> memory</i> , <i>send <someone> to the cleaners</i>
Covariational Conditional	The Xer the Yer (e.g. <i>the more you think about it, the less you understand</i>)
Ditransitive (double object)	Subj V Obj ₁ Obj ₂ (e.g. <i>he gave her a fish taco</i> ; <i>he baked her a muffin</i>)
Passive	Subj aux VPpp (PP _{by}) (e.g. <i>the armadillo was hit by a car</i>)

Despite their differing theoretical backgrounds, there is a considerable amount of overlap between CxG and corpus linguistics, and calls for an integration of these two approaches have thus been gaining traction recently, promoted by researchers such as Gries (2008), Groom (2017, 2019), Hunston and Su (2017), and Hunston (2019). Groom (2019) has noted that many corpus-based terms and concepts—such as Sinclair’s idiom principle—tend to be simply descriptive terms with no formal theoretical status. Meanwhile, the concept of constructions in CxG represents “a full and coherent theory of language as a symbolic system of usage-based form-meaning pairs” (311). Similarly, Hunston (2019), has proposed a framework in which corpus-based pattern grammar (Hunston & Francis, 1998, 2000), described in Section 2.4.3 below, can be applied to CxG for the purpose of identifying constructions. She argues that this integrated framework could

prove to be beneficial for teachers and learners since the construction grammar literature offers detailed descriptions of only a small number of constructions. By incorporating the currently available pattern grammar resources and reconfiguring them as potential constructions, a more comprehensive construction reference resource could be achieved.

2.3: Two Approaches to Phraseology

One of the main criticisms of phraseology has to do with the terminology associated with it: there seem to be nearly as many definitions of phraseology and phraseology-related terms as there are researchers in the field. However, two main categories have been identified, representing two different approaches to phraseology, which can help us make sense of the terminology. I have examined them and, in this study, will apply them as follows.

Granger and Paquot (2008, 28–29), borrowing terminology from Nesselhauf (2004), referred to a *phraseological approach* and a *frequency-based* (or *distributional*) approach. However, there is some risk of confusion when referring to “the phraseological approach” because it refers to an approach taken within the broader scope of phraseology. In order to avoid referring to “a phraseological approach to phraseology,” therefore, I will use Groom’s (2017) term *taxonomic approach* in place of *phraseological approach*. These are worth introducing in some detail.

The taxonomic approach, presenting another name for what Granger and Paquot introduced as the phraseological approach, is a deductive, top-down approach that is concerned with applying linguistic criteria and parameters to text. In this way, it identifies phraseological items, and it classifies these items in various ways. Examples of research following the taxonomic

approach include Carter's (1987), Gläser's (1998), Howarth's (1996), Cowie's (1998), and Mel'čuk's (2012).

The frequency-based approach, on the other hand, is not concerned with linguistic criteria or classifications, but rather uses a strictly corpus-driven, inductive methodology which relies on frequency and statistical analysis to identify recurring lexical phrases and patterns (see Sections 2.4.2 and 2.4.3). This approach, generally associated with the University of Birmingham and the work of Sinclair (e.g., 1991, 2003), produces a much wider range of lexical items, including many that do not fit neatly within the classification schemes of the taxonomic approach (Granger and Paquot 2008, 29). Other examples of this frequency-based approach to phraseology include studies by Stubbs (2002), Hanks (2005), Hoey (2005), Wray (2002), and Wikberg (2008).

I will further discuss these categories of terminology for my thesis in the remainder of this chapter section, across Subsections 2.3.1 and 2.3.2. I will then begin developing the appropriate terminology for my study, beginning with the foundation I will build on, in Subsection 2.3.3.

2.3.1: The Taxonomic Approach

Researchers following the taxonomic approach to phraseology are primarily concerned with identification and classification of phraseological items; Granger and Paquot (2008) cited Cowie (1981) as a typical example of such an approach. It is an approach in which phraseological items are placed on a continuum ranging from free combinations on one end to pure idioms on the other, with restricted collocations and figurative idioms in between these two extremes. According to Granger and Paquot (2008, 28–29), this approach is responsible for contributing to an established discipline of phraseology, and for creating terminology for the field.

2.3.2: The Frequency-based Approach

The frequency-based approach to phraseology is the newer of the two approaches, and it has been growing steadily in scope along with the fast-paced development of computing power and big data. Since this approach is inductive and data-driven, a much wider range of items can be identified as phraseological, beyond what would be considered within the scope of the lexical approach and its predefined linguistic categories. With the frequency-based approach, the growing field of phraseology has incorporated new concepts, such as collocational frameworks (Renouf and Sinclair 1991) and lexical patterns (Hunston and Francis 2000; see Section 3.4).

2.3.3: The Approach Taken in This Thesis

While Granger and Paquot (2008) acknowledged the seemingly irreconcilable differences between the taxonomic and frequency-based approaches to phraseology, they made a strong argument for an attempt at a merger.

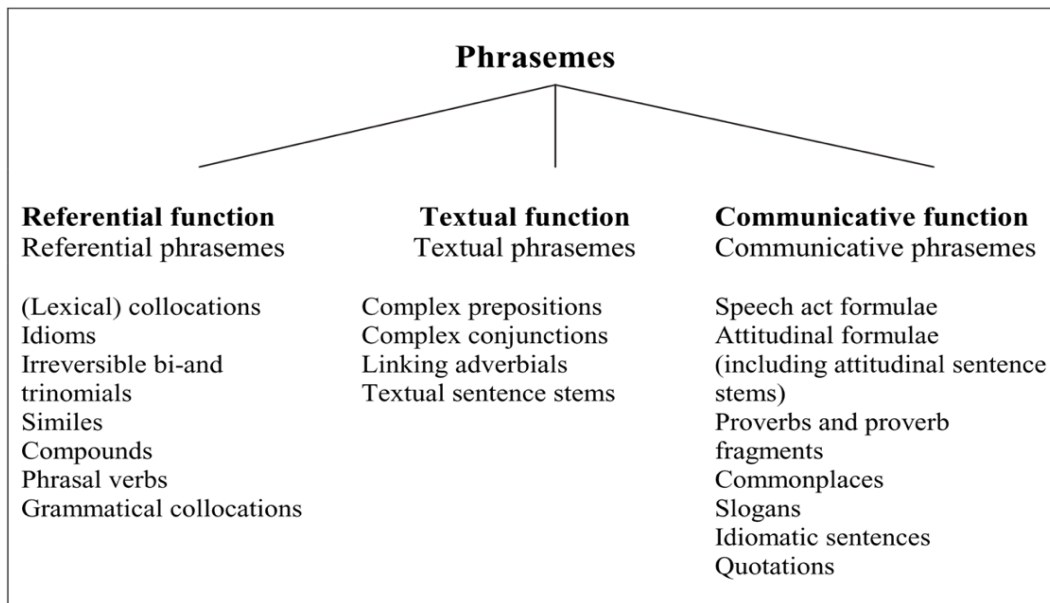
Proponents of the two approaches are still too wide apart and both sides have a great deal to gain from a rapprochement. Many linguists working in the traditional framework seem to be largely unaware of the benefit they could derive from automatic corpus-based methods of extraction and analysis. Conversely, linguists working in the distributional framework seem not to appreciate how much they stand to benefit from the fine-grained linguistic analyses of the traditional approach. (41)

Following this line of thinking, the methodology I will use in this thesis follows Walker's (2011) and involves elements of both the taxonomic and frequency-based approaches. All analyses will begin with a data-driven approach focusing on items around the node word *like*, with frequency and (when appropriate) statistical measures used as the main criteria for identifying lexicogrammatical phrases and patterns. Once identified, these items

will be analyzed from a taxonomic perspective to determine if they can be classified by type. This will be explained in more detail in Chapter 4.

While Granger and Paquot (2008, 41) supported merging the two approaches to phraseology, they warned that the terminology problem would need to be addressed, advising separate typologies for automated extraction techniques and linguistic analysis. Since my thesis will not involve automated extraction, only the linguistic analysis typology will be discussed in these pages, as demonstrated in Figure 2.1.

Figure 2.1: Granger and Paquot's phraseological spectrum. Reprinted from Granger and Paquot, [title], 2008.



As Figure 2.1 shows, Granger and Paquot (2008) proposed an extended version of Burger's classification scheme for phraseological items, consisting of three main categories: referential, textual, and communicative. This showed their attempt at laying foundations for a unified typology. The referential category is useful for classifying content messages (objects, phenomena, facts), while the textual is used for items structuring and organizing messages, and the communicative is used for expressions of feeling or

belief (or for directly addressing interlocutors). Granger and Paquot's classification scheme, with its three usage-based categories that allow for the addition of any phraseological items not initially included, is a good foundation. However, considering the enormous scope that phraseological studies now cover, it is inevitable that different researchers will continue to need slightly different terminology to discuss the specific items they are working with, and the present study, my thesis, is no exception.

2.4: Lexicogrammatical Units of Meaning

When dealing with phraseology-based terminology, I have found that the terminology problem discussed in Section 2.3 is present at all levels; it starts with the most general umbrella term used to describe all items. As seen in Figure 2.1, Granger and Paquot can be counted among several researchers using the term *phraseme*. Wray (2000) used the term *formulaic sequence* to collectively refer to idioms, collocations, and sentence frames, and Sinclair (2004) used the term *lexical item* to refer to either multi-word phrases or single words which function as discrete units of meaning. Because it seems to have a focus on phrases (or words grouped consecutively), I find *phraseme* to be a problematic term for my purpose in this study. However, phraseological research has shown that non-consecutive items, such as patterns and frames, are also crucial elements of the lexicogrammatical system. I will borrow the term *construction* from CxG, making it my main umbrella term for all the lexicogrammatical units of meaning that will be covered in this analysis of *like*. This will include discussions of collocations, phrases, and patterns, as well as more traditional grammatical constructions.

Viewing the lexicogrammatical system as a continuum ranging from the more grammatical and compositional to the more lexical, it is possible to talk about grammatical constructions and lexical constructions separately.

This is also true of the more lexicogrammatical constructions found between the two extremes. Each of these various kinds of construction will be discussed in Subsections 2.4.1 (collocation and colligation), 2.4.2 (phrase), and 2.4.3 (pattern and framework).

2.4.1: Collocation and Colligation

The terms collocation and colligation are attributed to Firth and his frequently cited assertion that “you shall know a word by the company it keeps” (1957, 11). Like much of the terminology associated with phraseology, these terms are used differently by different researchers, prompting Tucker (2006, 94) to label collocation a “thorny” topic. Generally, however, collocation is used in two main ways. First, when used as an uncountable noun, collocation refers to the process or tendency for certain words to frequently co-occur, with the co-occurrence being “more frequent than could be expected if words combined randomly in a language” (Nesselhauf 2005, 11–12). Second, when used as a countable noun, the term is used to refer to specific sequences of two or more frequently co-occurring words. In the present study, I will be using the term collocation only to refer to the process of co-occurrence. When referring to recurring instances of two or more words, I will either use construction (the more general term) or phrase (for more specific uses; see Subsection 2.4.2).

While collocation is typically used to refer to lexical co-occurrences, the related term *colligation* refers to grammatical co-occurrence. For example, in an analysis conducted by Sinclair (2004, 35), the author notes that the construction *true feelings* is a common collocation, where *true* and *feelings* are both lexical items which frequently co-occur with each other to form a single lexical item. Continuing his analysis, he finds that the collocation *true feelings* tends to strongly colligate with a possessive adjective, as in

“our true feelings.” Thus, colligation can be said to be a collocation featuring a grammatical element such as a specified part of speech or the passive voice.

When discussing collocation within a corpus-based framework, three items first named by Sinclair (1966) are typically used: *node*, *span*, and *collocate*. The *node*, in discussions of this kind, is the single word or phrase being investigated. *Span* refers to the number of words on either side of the node included in the analysis, and the *collocates* are those individual words found within the span that are said to collocate with the node. These terms help us visualize the process of collocation.

2.4.2: Phrase

The area of phraseology that has the most diverse range of terminology is related to recurring sequences of two or more collocating words. Table 2.1 provides an overview of some of the various terms used by different researchers, and their definitions.

As Table 2.2 shows, there is a great deal of variation in how these recurring sequences are defined, and this is of course a result of the different approaches, methodologies, and areas of interest of the respective researchers. Keeping this in mind, however, there is a commonality in that all of these definitions are meant to describe recurring sequences of words on the syntagmatic axis. For this thesis, in the name of simplicity, I decided to use the most generic of these terms: phrase. In this study, phrase will be defined as a construction that consists of a sequence of two or more words functioning as a complete lexicalized unit. In the data for this thesis, examples of phrases containing *like* include *stuff like that*, *that’s more like it*, and *sounds like fun*.

Table 2.2: Sample of terms used to describe recurring sequences of two or more words

Term	Researcher	Definition
Lexicalized sentence stem	Pawley and Syder (1983, 191)	A unit of clause length or longer with wholly or largely fixed grammatical form and lexical content
Lexical phrase	Nattinger and DeCarrico (1992, 1)	Conventionalised form–function composite that, compared to language that is put together each time, occurs more frequently and has more idiomatically determined meaning
Set phrase / phraseme	Mel'čuk (1998, 28)	A phrase which is <i>not free</i>
Fixed expression (including idiom; FEI)	Moon (1998)	Stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar
Formulaic sequence	Wray (2000, 465)	Linear sequence of two or more word forms occurring uninterrupted more than once in a corpus
Chain	Stubbs (2002, 230)	Any of the most frequent recurring lexical sequences in a register
Lexical bundle	Biber et al. (2004, 376)	Any of the most frequent sequences of words in a register
Lexical item	Sinclair (2004, 24)	A higher rank of lexical structure, above the word
Phraseologism	Gries (2008, 5)	The co-occurrence of a form, or a lemma of a lexical item and any other kind of linguistic element
Collocation	Walker (2011, 292)	A combination of two or more words which occur together or in close proximity to each other

2.4.3: Pattern and Framework

Most research in phraseology has kept a focus on collocation- and phrase-level constructions and their syntagmatic relations. More recently, however, the probabilistic approach has highlighted the importance of paradigmatic relations in a lexicogrammatical system. This is where *pattern* and

framework can be identified as two important types of construction, both featuring paradigmatic slots, as discussed in the literature.

For example, pattern has been defined as a construction of this kind. Particularly, it is a construction “frequently associated with (a sense of) a word, particularly in terms of the prepositions, groups, and clauses that follow [and precede] the word” (Hunston and Francis 2000, 3).

Nearer to emphasizing the grammatical in the lexicogrammatical, the term *framework* has been defined as consisting of one or more discontinuous sequences, with one or more lexical slots to be filled (Renouf and Sinclair 1991). With its paradigmatic slots, a collocational framework differs from a pattern in that patterns show connection to specific lexical items.

The definition of pattern being used here is one that Hunston and Francis (1998, 50–51) developed as they noticed that traditional functional categories (e.g., object, complement, or adjunct) could not adequately account for particular patterns of verb complementation. The authors began to form a simple system for notation, based on words and classes of words, that could transparently handle verb behaviours that had previously eluded description. Analyzing in this way, they clarified patterns found in traditional grammar prescriptions, familiar in ELT classrooms, such as *V to inf* (e.g., You have to weigh the responsibilities against the rewards) and *V —ing* (e.g., Snow began falling again. However, they also established patterns that had not appeared in pedagogic materials or in learners’ dictionaries, including *V about n* (e.g., I heard about the accident) and *V in n* (e.g., The books abound in social comedy).

This corpus-based pattern grammar lends support to Sinclair’s (1991, 53) assertion that sense and pattern are interrelated. For example, the two senses of the verb “recover” are reflected in their distinct differences of

pattern. One sense, “to get better,” works on the pattern of *V from n*, as in, “He is recovering from a knee injury.” This sense also works on a simple *V* pattern, given the example, “It took her three days to recover.” On the other hand, the other sense, “to get back,” shows the pattern *V n*, as found in “police . . . recovering stolen goods” (Hunston and Francis 1998, 51).

Hunston and Francis (2000, 3) reviewed this elaboration of pattern by noting that pattern depends on lexis and vice versa. A pattern shows up in relation to a particular set of lexical items, and each lexical item may appear in any of a set of patterns. Additionally, patterns are closely associated with meaning. This is first of all because the different senses of words can often be distinguished according to the different patterns in which they typically occur. Secondly, words that equally belong to use in a given pattern often show that they also share some aspect of meaning.

Related to the pattern grammar discussed by Hunston and Francis, Biber has proposed the term *lexical frame*. The main difference here is a question of method for identification. Gray and Biber (2013) identified, using automated extraction, the most common discontinuous sequences in a corpus. Meanwhile, the term *collocational framework* refers to another kind of lexicogrammatical construction with paradigmatic slots, one found more on the grammatical side of what we call lexicogrammatical (Renouf and Sinclair 1991). Patterns show connections with specific lexical items, but frames and frameworks are mainly discontinuous sequences in which lexical slots are open.

For the purposes of my thesis, pattern will be understood as a construction that consists of a lexical sequence which is known to recur, featuring at least one paradigmatic slot that is readily filled using any member from a set of lexical options. The set of lexical items that may fit the slot tend

to show grouping according to some kind of semantic set or grammatical class (Hunston and Francis 1998, 2000).

2.5: Identifying Lexical Items

As discussed in sections 2.3 and 2.4, the field of phraseology is mired by its enormous scope and lack of agreement among scholars with regard to terminology, and these unresolved issues can carry over to the realm of methodology. In this section, I will outline the various methods and rubrics used in identifying lexicogrammatical constructions.

Within the more traditional, taxonomic approach to phraseology, three main linguistic criteria are commonly used to identify lexical constructions: *lexicalization*, *fixedness*, and *non-compositionality*. Moon (1998, 6–8) provided a summary of these three criteria. First, lexicalization (or institutionalization) refers to “the process by which a string or formulation becomes recognized and accepted as a lexical item of the language” (Bauer 1983, 48). Moon states that lexicalization is a “necessary but not sufficient condition” for a sequence to be identified as a lexical construction, and she warns that because they are quantitative, “corpus derived statistics are no more representative than the corpus they relate to” (6). Next, Moon observes that fixedness involves “some degree of lexicogrammatical defectiveness in units” (7), including preferred lexical forms and grammatical restrictions. While Moon notes that corpus evidence does reveal certain levels of fixedness in lexical constructions, it is also apparent that variation is common. Finally, non-compositionality refers to the inability to determine the meaning of a lexical construction by combining the meaning of each individual word. It should be noted, however, that while researchers working within the framework of the probabilistic approach often consider lexicalization and

fixedness when identifying lexical items, semantic non-compositionality is not a necessary criterion (Groom 2017, 49).

In his examination of units of meaning, Sinclair (2004) outlined five categories of co-selection that form the components of a lexical item. He argued that the first two components, the *core* and *the semantic prosody*, are obligatory: the core of a lexical item “is invariable, and constitutes the evidence of the occurrence of the item as a whole” (141), and the semantic prosody determines the overall meaning of the lexical item. The remaining three components—collocation, colligation, and semantic preference—are optional and are responsible for “fine-tuning the meaning and giving semantic cohesion to the text as a whole” (Sinclair 2004, 141). Where collocation and colligation (see Section 3.4.2) refer to the co-occurrence of words and grammatical features respectively, semantic preference refers to the regular co-occurrence of items that share a common semantic feature.

Sinclair used the word *budge* to demonstrate this method of identifying a lexical item. After first noting that the treatment of *budge* in dictionaries is problematic, and carefully looking through a data set consisting of 31 concordance lines taken from a 20-million-word corpus, he provided the following analysis. The core of the identified lexical item is *NEG budge*, and it has a semantic prosody of frustration and irritation for someone who wants something moved. Next, he found that when the item that will not budge is an object, it frequently collocates with the lemma *refuse*, and has a semantic preference of refusal, generally through colligation with certain modals. This example shows how the traditional method of analysis at the single-word level can be problematic, since the dictionary definition that Sinclair first referred to did not mention the negative element attached to *budge*. This was true despite the fact that the examples provided by the dictionary had the

negative form. In Chapters 7 and 8 of this thesis, I will outline similar issues with *like*.

Gries (2008, 4) identified six parameters of phraseology necessary for identifying *phraseologisms*. He argued that these are important for researchers to carefully define when carrying out phraseological research, because they allow for comparison across studies without causing undue terminological, methodological, or theoretical confusion. Before continuing to the next section of this chapter, I will briefly outline and define these six parameters as they pertain to the current study.

(1) The nature of the elements involved in a phraseologism

As stated in 2.4, so as not to restrict or limit the possibility of identifying potential phraseological phenomena, I have chosen to define the phraseologisms examined in the current study rather broadly. I will be using *construction* as a term in the same sense as when used in CxG—as the most general, umbrella term to refer to any multi-word form-meaning pairing. More specifically, I will be using *collocation* as my term for the process of lexical co-occurrence. Similarly, *phrase* will be my term for referring to a continuous multi-word sequence treated as a single lexical item, and *pattern* will refer to a multi-word sequence with at least one paradigmatic slot to be filled.

(2) The number of elements involved in a phraseologism

With regard to phrases and patterns, I will not be placing any restrictions on the number of elements that can be involved in identified constructions containing *like*.

(3) The number of times an expression must be observed before it counts as a phraseologism

Generally, I will be using a cut-off frequency of 10 occurrences for any construction to be included in the final analyses of my study. I have chosen this number in order to limit the data to a more manageable level. The number is arbitrary of course, where some researchers may include in their research items with just two occurrences in a corpus.

- (4) The permissible distance between the elements involved in a phraseologism

Again, I will not be placing any specific restrictions on the distance between elements in a construction. However, in all collocational analyses carried out in this study, I will follow the standard four-word span, on both the left- and right-hand sides of the node, recommend by Sinclair. This span is also the default span used in the main collocation tool of the COCA.

- (5) The degree of lexical and syntactic flexibility of the elements involved

The only restriction that will be placed on this parameter will, of course, be the necessary inclusion of any form of the lemma *LIKE* within the construction. There will be no further restrictions placed on what type of lexical or syntactic elements are involved, allowing for full flexibility within the construction.

- (6) The role that semantic unity and semantic non-compositionality / non-predictability play in the definition

Gries (2008) noted that for many researchers, semantic unity is probably the most important parameter. He stated that “the elements of a phraseologism—however they are distributed across a clause or sentence—are usually assumed to function as a semantic unit, i.e., to have a sense just like a single morpheme or word” (6). This will also be the case in my study as presented in this thesis. However, non-compositionality, will not be required for a multi-word sequence to be considered as a construction.

2.6: Pedagogical Applications

2.6.1: The Lexical Syllabus

Over the years, numerous syllabus designs have been proposed corresponding to shifts in pedagogic trends and linguistic theories. In classifying these various syllabus types and their content, Nunan (1988) made the distinction between product-oriented and process-oriented syllabuses. A product-oriented syllabus focuses on the learner acquiring specific knowledge and skills, and a process-oriented syllabus focuses on the learning experience itself, with any given syllabus falling somewhere along a process/product continuum (Nunan 1988, 26).

Up until the 1970's, the majority of language syllabuses had been product-oriented, dominated by the structural-grammatical syllabus, which consists of a list of grammatical structures arranged in the order they are to be taught, based on frequency, perceived difficulty, usefulness, or any combination of these (Richards & Schmidt, 2002). With the emergence of communicative language teaching (CLT), however, several new approaches to syllabus design were developed including the process-oriented content-based syllabus (e.g. Mohan 1986) and the task-based syllabus (e.g. Willis 1996). As for product-oriented syllabuses, proponents of CLT began directing highly vocal criticism of structural-grammatical sequencing and its associated present, practice, and produce (PPP) methodology (Rutherford, 1987; Skehan, 1996; D. Willis, 1996). This criticism, in turn, led to the development of van Ek's (1975) threshold level English, and Wilkins' notional-functional syllabus, which focused on the communicative uses of language, as well as the lexical syllabus (Sinclair & Renouf 1988).

One of the main consequences of relying on a structural-grammatical syllabus is that vocabulary tends to be treated secondary in importance

(Sinclair and Renouf 1988, 141), and is often used unnaturally, with its main purpose serving to reinforce the more centrally emphasized grammatical concepts. This composition, however, arises from that traditional slot-and-filler belief where grammar and lexis are held independent of one other. As an alternative to this structural approach, the lexical syllabus was first proposed by Sinclair and Renouf (1988) and more fully explored by Willis and Willis (1988), Willis (1990), Nattinger and DeCarrico (1992) and Lewis (1996). A lexical syllabus may be defined as follows:

A vocabulary syllabus that is organized in terms of the most important, frequent, or useful vocabulary items in a language. Lexical syllabuses are often organized according to levels (e.g. the first 1000 words, the second 1000 words, etc.). (Richards and Schmidt 2002, 307)

Some scholars (e.g., McCarthy 1990) have claimed that it would make sense, in teaching a language, to begin by presenting the most frequent words of that language. Sinclair and Renouf (1988) argued that language learners' main focus should be not only those most common words, but their most common forms, with their main patterns of use and most frequent combinations (148). They noted that, surprisingly often, a lemma's different forms may import different meanings. They further claimed that these different forms were basically different words, and that these new versions of the words should be presented as distinct elements for the purposes of designing language courses (147). They cited evidence from the Birmingham Corpus (now known as the Bank of English) to demonstrate that word pairs including "certain/certainly," "near/nearly," and "detach/detached" behave differently despite their common semiotics (Sinclair and Renouf 1988, 147–148). They also observed that, without taking a lexical approach, authors of textbooks could not place suitable focus on the delexical verbs (such as "have" or "give,") in their tendencies to collocate with given

nouns and adjectives, thus forming items that are used with greater frequency than the more concrete lexical verbs available in each case (153).

Willis and Willis developed their *Collins COBUILD English Course* (CCEC) from the Birmingham Corpus, which consisted at that time of some 20 million words. The lexical syllabus for this course focused on the fact that 70% of English texts are composed of just 700 words from the top of the corpus, ranked by frequency (Willis and Willis 1988, 5). The course met with praise (on the part of some reviewers) for its approach, but its reception was also marked by complaints (on the part of other reviewers) for being difficult to use in the classroom (see Taylor 1991). Sinclair and Renouf (1988) had predicted such mixed reactions, admitting there would be a radical sense of difference about a lexical syllabus as compared to conventional textbooks, and that it would “certainly meet resistance at first” (160). However, when asked why he thought the lexical syllabus and CCEC were not as successful as might have been expected, Willis (personal communication, 2009) noted, in addition to the issue of being the first with a new innovation which takes time to catch on, that there may have been issues with the marketing strategy, or lack thereof, for the lexical syllabus and CCEC at the time. Perhaps more importantly, Willis took issue with Lewis’ Lexical Approach, stating as follows:

One of the problems with the lexical syllabus was the success of Michael Lewis’s Lexical Approach. Lewis misrepresented work done on the lexical syllabus by saying that it was word based and ignored the importance of lexical phrases. What we (Sinclair and Renouf, 1988, Willis 1990) actually said was that the syllabus should be made up of the most frequent words with their most frequent meanings and patterns of occurrence, and that the most efficient way to identify phrases was by starting from the most frequent words. Lewis emphasised lexical phrases but ignored the importance of words as a starting point. So Lewis’s lexical approach had a beneficial effect on syllabus design and teaching in that it

highlighted the importance of lexical phrases. But in shifting the emphasis from the most frequent words Lewis threw the baby out with the bathwater. He recognised the central importance of lexical phrases, but rejected the means by which phrases could best be identified as a starting point for the process of syllabus design.

It seems then, that based on Willis' statement, the lexical syllabus, unfortunately, had several problems which hindered a more widespread adoption by both educators and ELT publishers.

2.6.2: Consciousness-Raising and Data-Driven Learning

The traditional structural-grammatical and functional syllabuses often found in ELT coursebooks are often realized through Presentation, Practice, and Production (PPP) methodology. As classroom methodology, PPP generally consists of explicitly presenting grammar rules and vocabulary lists that are to be memorized. This deductive teaching approach has held on despite evidence that suggests learning is more effective when students are encouraged to analyze data and formulate hypotheses for themselves (Sharwood-Smith, 1981; Rutherford, 1987; Herron & Tomasello, 1992; Boulton, 2007a). This latter process, which is termed consciousness-raising (CR) by Rutherford (1987), consists of using inductive teaching strategies which help raise students' awareness of specific features of the target language—without the explicit presentation of rules. In this way, language learning is considered more effective within a process-oriented syllabus than a product-oriented syllabus (Nunan, 1988).

Developed by Johns (1991, 2) on the premise that “research is too serious to be left to the researchers,” data-driven learning (DDL) raises language learners' consciousness by presenting corpus concordance lines, raw or pre-selected. The method allows learners to draw conclusions on their own as they observe subtle differences in usage language while they focus on the target language (p. 2). This method fits well with questions from

students about the difference between one item and another, or about which of two or more options is more “natural,” leading the teacher and students to look at raw corpus data together. The answers are not predetermined; the students’ hypotheses can lead them to formulate their own rules, informing a complete classroom pedagogy. Furthermore, DDL calls for a certain amount of data manipulation on the teacher’s part when the focus is to be on concordance lines rather than raw data. By carefully selecting (and occasionally editing) these lines the teacher can adapt the method for lower-level students. This greater degree of teacher control can be useful in the exploration of targeted lexicogrammatical patterns, especially those considered either particularly useful or particularly difficult for the students (Johns 1991, 2; Hunston and Laviosa 2001, 129).

Figure 2.2 presents a DDL exercise from Johns (1994). Here, data are manipulated, and students are led, through a predetermined series of notes based on inductive reasoning. Verbs and nouns appearing before the clause containing *that* show a similarity of meaning in each concordance line, and they specifically inform similarities in patterns of functional meaning, including “N *that* SVO” and “V *that* SVO.” Note that in these examples N = “proposal/recommendation/suggestion,” and V = “propose/recommend/suggest”. “Should” is omitted from the second group of citations, and the rule that may be observed is that “should” may be omitted if redundant with regard to either the verb or the noun in the main clause.

Figure 2.2: *That-* clauses DDL exercise. Reprinted from Johns, in *Perspectives on Pedagogical Grammar*, edited by T. Odlin, 1994.

<p>Look carefully at the following citations from <i>New Scientist</i>. What two features do they have in common?</p> <p>1) ena from more dubious data, we propose that they should be renamed 'UAPs', for unidentified</p> <p>2) hen his advisory committee recommended that last year Depo-Provera should get a license. C</p> <p>3) eir drinking water. The EEC recommends that tap water should not contain more than 5 micro</p> <p>4) he scale of the problem. It recommends that the government should set a firm date for thes</p> <p>5) olol. The Greenfield report recommends that prescription forms should contain a box, which</p> <p>6) the road plan was the recommendation that for half its length it should be routed through</p> <p>7) h a name for the strategy. He suggests that the term 'porpoising' should be used to descry</p> <p>8) put an absolute veto on-any suggestion that Post Office canvassers should be remunerated b</p>
<p>Now look at the following citations. How are they similar to citations 1–8? And how are they different? Can you explain the difference?</p> <p>9) sed, and with it, speed. They proposed that Harvard create such a super-track to hum</p> <p>10) ed voice: 'Then will somebody propose that this paper be rejected irrespective of its co</p> <p>11) under review. An HSE document proposes that GMAG be turned into ACGM - an Advisory Comm</p> <p>12) apolis. They said: '. . . we recommend that the dose of benoxaprofen be decreased (approx</p> <p>13) mistic. The committee also recommended that the government clarify the rules covering the</p> <p>14) sing plants breaks down. It recommends that France take a second look at following the pol</p> <p>15) h. The coordinating committee suggests that the appeal panel ask why this change has been</p> <p>16) of the University of Bristol, suggests that a group of babies be trained to use a 'baby-op</p>
<p>Who can think of the most interesting completions for the following sentences?</p> <p>1) If I were head of my department I would recommend that all examinations_____.</p> <p>2) When John told me that he had an argument with his girl-friend I suggested that he _____.</p> <p>3) As a postgraduate student at the University of Birmingham, I propose that _____.</p>

As a teaching methodology, DDL is suitable for student-centred classrooms with a focus on lexis; its results have been verified through both quantitative and qualitative evaluation (e.g. Johns 1991, 1994; Tian 2004; Bolton 2007a, 2007b, 2008; Peppard 2014; Cobb and Boulton 2015). In one such experiment, Tian (2004) found that students using DDL showed significantly greater improvement as compared with a control group. The control group had received instruction using a traditional, deductive presentation style, and the DDL group outperformed the control group in word use and in understanding text type features.

However, certain drawbacks are likely to hinder DDL's acceptance and growth among mainstream classroom methodologies. For example, worksheets for DDL often take considerable time to prepare, especially when they need to manipulate data as in the example shown in Table 2.2. Johns (1991, 4) reported preparation time taking anywhere from 40 minutes to four hours. Many teachers, perhaps most today, at private language schools and in wider school systems alike, lack significant amounts of time for such tasks on a daily basis. Moreover, access to a corpus may not be easy to obtain, and it can take some time for a teacher to become familiar with the technology. Additionally, Johns (1991, 3) admitted, some teachers might not be willing to relinquish control of the classroom activities in allowing students to initiate research. This raises the question of whether (and how) DDL might be ready for use in relation to more traditional methods.

2.6.3: Implications for Coursebook Syllabus Design

It has been more than 30 years since Sinclair and Renouf's (1988) first proposal of a lexical syllabus, and few notable applications have appeared in commercial distribution, suggesting that resistance has been strong enough to lead to a rejection. The norm remains in favour of an accumulated entities view (see Rutherford 1987), in which language learning is to be realized through grammar-focused instruction in a PPP format, one item at a time (Criado 2013). Research in second language acquisition is less hesitant to refute the validity of this model, but it is hardly surprising in light of Littlejohn's (1992) investigation of materials for the EFL classroom. Given the potential for research in applied linguistics to inform coursebooks, this study found a relatively weak influence. Chalker (1994, 41) did not find significant differences in the contents and organization of coursebooks from the structuralism-dominated 1950s and 1960s, compared with those of the 1990s—

when communicative language teaching had been established as a better model than structuralism. Despite this perceived shift away from structural-grammatical sequencing in favor of more communicative-based approaches, it seems that the majority of modern ELT coursebooks are still based on a structural-grammatical syllabus, or organized around a grammatical syllabus while styling themselves as thematic and communicative (Martel 2013, 1122). As for the potential of corpus linguistics, O’Keeffe et al. (2007) reported that there was a “frequent mismatch between corpus linguistics research and what goes into materials and resources, and what goes on in the language classroom” (xi).

There are several probable reasons for the structural-grammatical syllabus and PPP’s persistence:

- (1) It affords a degree of control throughout a lesson, comforting for both teachers and syllabus writers; it is easily planned, beginning by clearly specifying goals and similar objectives that can easily be tested (Skehan 1996, 17).
- (2) Its general methodology, like its particular techniques, are easy to adapt for teacher training (Skehan 1996, 17)—and this is a valid concern for the profession of teaching because of its high rate of turnover.
- (3) It lends to learners’ beliefs and opinions about what their teachers should do, as long as it remains a dominant approach and therefore part of learners’ previous experiences and expectations based on cultural influences (Lightbrown and Spada 2006, 67).
- (4) It is the basis for large volumes of sales revenue for publishers in ELT, whose reluctance to deviate from what works to their bottom-line advantage underscores a lack of demand for alternatives that could be based on more recent research; publishers may be averse to the risk of adapting their products to reflect such research.

It therefore appears that changing how coursebooks are made will require a subtle approach. As mentioned above, the *Collins COBUILD English*

Course (Willis and Willis 1988), for example, was published but hardly accepted among teachers or their students. A compromise should nevertheless be possible. It is my belief that this compromise can be achieved by incorporating a lexical syllabus realized through integrated DDL exercises within an otherwise “traditional-looking” coursebook (see Peppard 2014).

2.7: Conclusion

In this chapter, I have explored the topic of phraseology in order to provide the background and framework for the study I will present in this thesis. In Section 2.2, I have examined several theories that support a phraseological view of language, contrasting these with the more traditional, generative, view of language that treats syntax and lexis as separate entities. In Section 2.3, I have outlined two different approaches to researching phraseology, providing examples from the literature to illustrate these two approaches. In Section 2.4, I have examined phraseological units of meaning, addressing the problem of terminology in the literature, with detailed explanations of the concepts and definitions to be used in this study, of collocation, lexical phrases, and lexical patterning. In Section 2.5, I have outlined the various ways I will use for identifying lexical items. And, in Section 2.6, I have looked at two complimentary pedagogic applications of corpus-based research in phraseology, providing examples of how they can be applied to the kind of research I will carry out in this thesis.

3

Review of Literature

Simile

3.1: Introduction

Prepositional *like* is highly frequent in both spoken and written English, due in large part to its role as one of the main function words used to form similes. While the language of simile has been examined rather extensively, as evidenced in the literature, it is rarely the main focus of study. This is due to the majority of research on simile being concerned with the relationship between simile (*A is like B*) and metaphor (*A is B*), with the former often treated as a subset, or lower form, of the latter (Hanks 2005,1; Littlemore and Low 2006, 3; Veale and Hao 2007, 683). Fadaee (2011, 22), acknowledging this incongruence, pointed out that simile is investigated much less than metaphor despite occurring just as frequently in discourse. Furthermore, there are some phraseological studies of simile in the literature, but it appears that there is a lack of corpus-based research, since many studies suffer from the limitation of using simplified examples that have been removed from their context (Wikberg 2008, 129). There are even fewer corpus-based studies of simile that examine genre differences in the usage of simile.

In this chapter, I will provide a thorough review of the existing literature on simile in order to establish the framework for the current study. By providing a corpus-based analysis of simile usage with *like* that focuses on

both phraseology and genre differences, I will also attempt to fill some of the perceived gaps in the literature.

In Section 3.2, I will operationalize the definition of simile that is to be used in this thesis. And, by collating the various definitions found in the literature, I will provide a four-point framework for identifying similes in corpus data, paying special attention to the distinction between literal comparison and true simile usage. This will be followed by an explanation of the supporting terminology used to describe the elements of a simile. Section 3.3 will review the existing literature, comparing simile to metaphor, and it will provide evidence for the position taken in this thesis: that simile should be treated as a distinct element of figurative language, one that is independent of metaphor. In Section 3.4, I will provide a brief outline of the various types of simile that have been identified in the literature, with more thorough discussions of the types relevant to my thesis, as they will appear in the results (see Chapters 5 and 6). Section 3.5 will provide a review of the various ways used in the literature to categorize and analyze simile, with special focus on previous phraseological and corpus-based studies. Finally, in Section 3.6, I will outline the research on genre differences and simile usage; this will provide the springboard for a significant portion of the current study.

3.2: Defining Simile

While the concept of simile is relatively simple and well-known, many of the available definitions of simile found in the literature are inadequate for the purposes of this thesis. Some dictionary definitions, for example, are not specific enough as we can see in the following from the website for dictionary.com (accessed September 6, 2018): “[a simile is] a figure of speech in which two unlike things are explicitly compared, as in ‘she is like a rose.’”. This definition is problematic because it makes no mention of the need for

a function word such as *like* or *as*, and therefore fails to make the distinction between simile and metaphor. Wikberg, citing Miller's (1993, 373) definition in which he describes simile as "a comparison statement involving two unlike things," noted that this inadequate sort of definition is found not only in dictionaries, but also in academic research.

Richards and Schmidt (2002, 201), however, provide a useful definition of simile in the academically-oriented *Longman Dictionary of Language Teaching and Applied Linguistics*, found under the entry for *figure of speech*. Here, the authors identify simile and metaphor as being the two most common types of figure of speech

A simile is an expression in which something is compared to something else by the use of a FUNCTION WORD such as *like* or *as*. In *Tom eats like a horse*, Tom's appetite is compared to that of a horse. *My hands are as cold as ice* means that my hands are very cold.

This definition is adequate for general reference and covers the main elements of what makes a simile a simile. However, it does not specify what type of comparison is being made, and it fails to provide the necessary tools for separating true similes from literal comparisons.

Due to this lack of a readily available and fully comprehensive definition of simile suitable for the present research, I have collated from the literature the following four-point framework for defining and identifying true similes:

- (1) A simile is a comparison between two unlike things (Miller 1993, 373).
- (2) A simile is an assertion of similarity (Ortony 1979, 162).
- (3) Similes are explicitly marked with a function word such as *like* or *as* (including *as if*, *as though* phrases, etc.; Fishelov 1993, 8).
- (4) Similes are comprised of figurative comparisons, i.e., they are not literal; this can be tested by means of a symmetry test, where true similes are not symmetrical (e.g., $A = B$ and $B = A$ in exactly the same way; Ortony 1979, 171).

In the sections below, a more detailed discussion of literal versus figurative comparisons, and of using symmetry tests to identify similes, will follow a brief description of the terminology used for talking about similes.

3.2.1: Simile Components

A typical simile is comprised of four main elements: the target, a function word, the vehicle, and the grounds. The *target* of the simile is the object, entity, state, or process that is being described in the comparison. Other terminology is found in the literature to refer to the target of a simile, including *topic* (Ortony 1979; Wikberg) and *tenor*, the latter introduced by Richards (1936) to identify the elements of a metaphor. The term *target* is preferred here, and I will use it throughout this study. Target has the inferred meaning of goal; when one uses a simile, they have a goal of communicating one thing in terms of another. The *vehicle* of a simile refers to the object, entity, state, or process used to make the comparison in describing the target. The *grounds* of the simile refer to the similarities (between the target and the vehicle) that establish the simile. Finally, the function word is the defining element of the simile that marks the comparison. The most common function words used to form similes are *like* and *as*, but the phrasal conjunctions *as if* and *as though*, and the lexical frame *as ADJECTIVE as* are also used.

Bridgeman (1996, 67) describes the *BE like* syntax of a simile as a guarantee that the intended similarity between the target and the vehicle will be discoverable. While *like* and *as* are the most commonly used (and most mentioned) function words, there is a very large number of expressions that can be used, including less direct phrases such as *seem to* and *recall* (Fishelov 1993, 8).

Similes can be either closed or open, which refers to the presence or absence of explicitly stated grounds (Margolis 1957, 186).

- (3.1) This is the time that I shine bright **like** a diamond. (NEWS[WashPost])
- (3.2) Each song was **like** a diamond, and my first thought was, "damn he's good."
(SPOK[CBS_Morning])

Example 3.1 is a closed simile. The target, "I" is being described by making a comparison to the vehicle "a diamond" with the function word "like." The simile is said to be closed because the grounds of the comparison are explicitly stated, in that both the target and the vehicle "shine bright." In Example 3.2, the target ("each song") is compared to the vehicle (again "a diamond") via the function word "like," but here the grounds are not given, leaving interpretation to the listener. While it can be said that open similes are less determinate and more susceptible than closed similes to differing interpretations, they are not necessarily more difficult to understand (Fishelov 1993, 8). In Example 3.2, the intended meaning is quite clear. The speaker is using the lexicalized, metaphorical meaning of diamond to describe the songs, that is, they are beautiful, brilliant, precious, and so forth.

3.2.2: Simile and Literal Comparison

While all similes are comparisons, the reverse is not true (Bredin 1998, 7). In determining whether a comparison is figurative or literal, it is often necessary to perform a symmetry test (Ortony 1979, 171), or what Wikberg referred to as a reversibility test. This is because true similes are not symmetrical or reversible (i.e., it is not true that $A = B$ and $B = A$ in exactly the same way).

- (3.3) The race car is **like** a fighter aircraft. It's built for one purpose only, to go the fastest, be the fastest, no nonsense, just one purpose, like

a fighter in combat only. No comfort of any sort,
but has a limit. To be able to operate, take
advantage of what you have, you have to be able
to--just ride that limit. (SPOK[ABC_Nightline])

In Example 3.3, *race car* and *fighter aircraft* are reversible, and reversing them does not alter the meaning of the comparison, hence it is a literal comparison and not a true simile. As noted by Wikberg (130), literal comparison is made possible because of the two elements of the comparison being from very similar categories—in this case, highly specialized machines. Ortony (1979) referred to this as *high salience*. Additionally, literal comparisons often necessitate some elaboration or explanation of the similarities, a phenomenon that is clearly illustrated in the example given above.

However, full explanation of the similarity in a *like* comparison does not specifically identify a literal comparison. Roncero, Kennedy, and Smyth (2006) found that similes on the Internet often contain explicit explanations of the grounds, and these explanations of grounds can also be very elaborate. This is exemplified in the following example, from the Magazine section of the Corpus of Contemporary American English (COCA):

(3.4) Kitchen gadgets are **like** toys for grown-ups:
Finding the right one can turn rushed, in-a-rut
dinner prep into a healthy adventure. Once you
see how a cleverly designed grater or nonstick
pan can transform your usual meal into a
nutrient-rich sensation, you'll be inspired to
play and make your own delicious creations.
(MAG[Prevention])

In Example 3.4, reversing the target and vehicle, “Toys for grown-ups are like kitchen gadgets,” does not work and the purpose of the comparison is lost. This, along with Example 3.3, demonstrates the necessity of the reversibility test and consideration of target–vehicle salience when identifying similes. This can be a time-consuming task when working with large

amounts of corpus data, and as Wikberg (140) noted, despite the easily identifiable syntactic properties of similes, “careful scrutiny” of the co-text is often necessary for positive identification.

3.3: Simile and Metaphor

Much has been written about the similarities and differences between simile and metaphor. At a base syntactic level, a simile is an overt comparison, while the comparison formed in a metaphor is covert (Fadaee 2011, 22). In discussions of figurative meaning, it has traditionally been thought that simile expresses the same type of meaning as metaphor despite their differing sentence patterns (Toyomura et al. 2012). Indeed, the two figures of speech have been treated as equal by many writers since the days of Aristotle (Glucksberg and Haught 2006, 360). The most recent research, however, has provided evidence suggesting that simile and metaphor are distinctly independent of each other. For example, Littlemore and Low (2006, 43) point out that while similes generally focus attention on typical, or central, characteristics of the vehicle, metaphors can involve more peripheral, or non-central, characteristics, which necessitates more conceptual restructuring to interpret. Currently, a large part of the literature involving both simile and metaphor stems from debate concerned with three competing psycholinguistic models of metaphor. These will be outlined in the following sections.

3.3.1: Three Psycholinguistic Models

Psycholinguistic views of metaphorical language, as it is produced and comprehended, have traditionally fallen into two competing categories, known as the comparison model and the categorization model (Glucksberg and Haught 2006, 360), or the equivalence view and the nonequivalence view

(Aisenman 1999). A third model, the career of metaphor model, proposed by Bowdle and Gentner (2005) aims to bridge the gap between the two main theories.

3.3.1.1: The Comparison Model

Proponents of the traditional comparison model of metaphor (e.g., Miller 1993; Ortony, 1979; Johnson and Malgady 1979) essentially see no difference between simile and metaphor. This view dates back to Aristotle's *Rhetoric* (Aristotle 1984), in which the philosopher wrote as follows:

The simile also is a metaphor; the difference is but slight. When the poet says of Achilles: "He leapt on the foe as a lion," this is a simile; when he says of him "the lion leapt," it is a metaphor—here, since both are courageous, he has transferred to Achilles the name of "lion." Similes are useful in prose as well as in verse; but not often, since they are of the nature of poetry. They are to be employed just as metaphors are employed, since they are really the same thing except for the difference mentioned. (2,243)

According to this view then, a metaphor is simply an abbreviated simile (Shibata 2012, 101). In the example provided by Glucksberg and Haught (2006, 361–362) the simile, "My lawyer is like a shark," would be considered equal to its corresponding metaphor: "My lawyer is a shark."

A more recent, and well-known, conceptual approach to metaphor, put forward by Lakoff and Johnson (1980, 2003), falls under the comparison view of metaphor. Lakoff and Johnson treat metaphor as a conceptual phenomenon rather than a linguistic one, where a mapping process occurs between a base and a target domain, thus making a comparison. If the difference between simile and metaphor is most notably a question of whether a function word such as *like* is used or not, this would be considered a superficial linguistic difference. Because of this, both figures of speech are treated

as if they were the same “with regard to comprehension, interpretation, and usage” (Aisenman 1999, 46).

When first introduced, the conceptual view of metaphor was well received and, as Steen (2008, 215) noted, it “triggered a revolution in metaphor studies” and found support from psycholinguistic research such as that presented by Gibbs (1994). More recently, however, researchers have begun identifying problems with this theory (McGlone 2007) and, most relevant to the present thesis, there is a growing body of evidence suggesting the differences between metaphor and simile are not at all linguistically superficial.

3.3.1.2: The Categorization Model

The categorization view of metaphor rejects the idea that metaphor and simile are interchangeable as claimed in the comparison model. Instead, in this view it is argued that the two figures of speech rely on different cognitive processes for comprehension, where simile is processed as a comparison and metaphor is processed as categorization (Glucksberg 1980; Glucksberg and Keysar 1990; Glucksberg and Haught 2006b).

Returning to Glucksberg and Haught’s (2006) example, in the simile, “My lawyer is like a shark,” the categorization view of simile assumes that “shark” refers to the predatory fish. However, in the corresponding metaphor, “My lawyer is a shark,” the “shark” refers to a metaphorical representation of sharks as predators, and thus the lawyer is being categorized as a predator.

A growing body of evidence now supports the validity of the categorization model of metaphor over the comparison model, with the most notable being that presented by Glucksberg and others (Glucksberg and Haught 2006a, 2006b; Glucksberg and Keysar 1990). Through a series of

experiments reported in their papers, they concluded that “because a metaphor cannot always be understood in terms of its corresponding simile, . . . comparison theories of metaphor are fundamentally flawed” (Glucksberg and Haught 2006, 360). A good example of how they are fundamentally flawed can be found in Hanks’s (4–5) corpus-based research on the preposition *like*, in which the author showed how certain types of simile cannot be successfully translated into metaphor. For example, there is the case of *VERB like* similes formed with a verb of perception and an irrealis vehicle (a vehicle based on an imagined and unreal entity), such as, “I took the pastry; it tasted like sweetened cardboard.” This clearly differs from saying that the pastry was sweetened cardboard.

Research conducted by Aisenman (1999) into the functional differences between simile and metaphor adds further support for the categorization model. Based on two preference task studies, Aisenman found that similes are preferred for mapping attributive predicates (on the pattern, “A is B”) and metaphors are favoured for mapping relational predicates (such as “A contains B”). More recent psycholinguistic research, conducted by Shibata (2012) and Shibata et al. (2012), offered further evidence for a distinction between simile and metaphor. Using functional magnetic resonance imaging, these researchers found that higher levels of activation were observed in the medial frontal region of the brain for similes and in the right-sided prefrontal region for metaphors. (Both simile and metaphor processing were known to occur in the left frontal region.) They suggested that the area active for comprehending similes could be related to inference processes, while the area active for processing metaphors might be related to comprehending figurative language.

While Glucksberg (Glucksberg and McGlone 2001) at first wholly rejected the comparison view of metaphor processing, including Lakoff and Johnson's (1980) theory, he later came to a more conciliatory approach. He admitted that it was possible for metaphors to be processed comparatively, and now posed the question: "when and under what circumstances are metaphors processed as categorizations, and when as comparisons?" (Glucksberg and Haught 2006, 362)

3.3.1.3 The Career of Metaphor Model

The career of metaphor model (Gentner and Bowdle 2001; Bowdle and Gentner 2005) attempts to answer the question posed by Glucksberg and Haught (2006) while bridging the gap between the opposing comparison and categorization views of metaphor. It proposes that novel metaphors are processed comparatively: generally, they are processed as similes but, over time and repeated use and exposure, as they become conventionalized each can be processed as forming either a comparison or a categorization.

Evidence that can be found in the literature both supports and refutes the career of metaphor hypothesis. Several studies (Chiappe and Kennedy 1999; Chiappe, Kennedy, and Chiappe 2003; Jones and Estes 2005) have suggested that aptness, rather than conventionality, is the primary determiner of metaphorical processing. These findings have found support in a series of experiments carried out by Jones and Estes (2006) finding that aptness predicted a preference for similes over their corresponding metaphors. It also predicted the speed and ease of metaphor comprehension, and the categorization of metaphorical terms. On the other hand, however, in a series of three experiments comparing conventional and novel statements with aptness ratings, Gokcesu (2009, 571–572) concluded that conventional figurative statements seem to be processed categorically while

novel figurative statements seem to be processed comparatively. This provided evidence in support of the career of metaphor model. It appears that more research needs to be carried out on the career of metaphor hypothesis before a definitive conclusion can be formed as to its validity.

3.4: The Approach Taken in This Thesis

I will not take a stand on the relative merits of the categorization and career of metaphor models of metaphor and cognitive processing, as this would lie far outside the scope of the research for this thesis. However, it does seem that, based on the supporting evidence from both of these models as outlined in 3.2, we can safely conclude that simile and metaphor are inherently different. I have conducted the research for this thesis under the assumption that simile is a distinct linguistic phenomenon and not simply a lower form of metaphor. And, for my purposes, simile is indeed a worthy subject of investigation in and of itself, without the need to connect it to metaphor.

3.5: Types of Simile

3.5.1: Classification Schemes

A review of the literature reveals that there is no agreed-upon method of categorizing simile and metaphor types, as several very different categorization schemes exist based on differing angles of analysis. In addition to the open simile / closed simile dichotomy outlined previously, Ortony (1979) utilized a simple classification of literal and non-literal comparison statements in his analysis. Expanding on this, Fishelov (1993), in addition to literal comparisons, distinguished between poetic and non-poetic similes. In yet another classification scheme, Fromilhague (1995) made a distinction between objective simile and subjective simile, with the former originating

from concrete physical experience and the latter from individual association mechanisms. He also distinguished between explicit simile (where the grounds are stated directly) and implicit simile (where the grounds are not stated, leaving the interpretation up to the reader or listener). This distinction, too, corresponds to open and closed simile as described by Margolis (1957), which highlights the problematic lack of agreement concerning accepted terminology in this field.

Bredin (1998, 69) identified six types of logical comparison statements and argued that these translate into six corresponding types of simile:

- (1) A is like B.
- (2) A is unlike B.
- (3) A is like B in respect of p.
- (4) A is unlike B in respect of q.
- (5) A has as much as r as B has.
- (6) A has a different quantity of s than B has.

Cameron and Deignan (2003) take a very different approach to identifying and classifying metaphor use with their notion of *tuning devices*, which they define as expressions used to help the interlocuter interpret the intended meaning of a metaphor. In their research with both large and small corpora, they found that tuning devices frequently occur with metaphors in the data and are used for various discourse “tuning” functions, “such as alerting the hearer to any problem in interpretation, and suggesting which interpretation, literal or metaphorical, is intended” (150). Cameron and Deignan’s (2003) work falls within the comparative model of metaphor, and they therefore make no distinction between simile and metaphor. Therefore, in their analysis, items such as *actually*, *imagine*, *just*, *kind of*, and *like* are viewed as tuning devices, or what Goatly (1997) refers to as signalling devices, used alongside metaphors. In the case of *like* usage, rather than

being used as a simile marker, *like* is viewed as a tuning device used to indicate the nature of the topic-vehicle mapping (153).

Finally, basing her analysis on grammatical structure, Wikberg noted that the vehicles of similes fall into three main categories: predicative [1], adverbial [2], and clausal [3]. Predicative similes can be further divided into nominal [1a], verbal [1b], and comparative [1c] forms, and adverbial similes can be further divided into nominal [2a] and comparative [2b] forms. In addition, two subcategories of verbs are used in adverbial similes: the predicative subcategory [2a¹] consists of the verbs of perception (LOOK, SOUND, FEEL, TASTE, SMELL) and SEEM. And, meanwhile, the adverbial subcategory [2a²] consists of a wide range of verbs other than *BE* and the perception verbs, such as BEHAVE, ACT, and DRIVE.

- [1a] Her voice was **like** slow thunder and sweet rain.
(FIC[Storyworks])
- [1b] It was like being **stoned** to death with popcorn.
(MAG[USCatholic])
- [1c] He's as strong as a bull. (MAG[SportsIll])
- [2a¹] This looks like - it looks **like** crap.
(SPOK[CNN:ShowbizTonight])
- [2a²] Humanity would explode **like** a nova.
(FIC[BkSF:WhenFiveMoons])
- [2b] Indeed, the place is run as tight as a nuclear submarine. (MAG[Inc.])
- [3] It's **like** all my CVs are going into a black hole.
(ACAD[CommunCare])

The study for this thesis will start with this grammatical classification of similes at its base, as it is the most objective and amenable to a corpus-based approach. Considering that the focus of this research centres around similes formed with *like*, comparative *as . . . as* similes will not be included in the analyses.

3.5.2: Lexicalized and Innovative Similes

Another distinction that can be made with regard to simile use is whether a simile is lexicalized (also referred to as phrasal [Wikberg], and conventional [Hanks]) or innovative. As stated in Section 2.5, Moon (1998, 6–8) outlined three criteria to take into account for the identification of lexicalized phrases: institutionalization, fixedness, and non-compositionality. These will be briefly revisited below.

Institutionalization refers to the process that occurs when a recurring phrase becomes accepted and recognized by the language community as a lexical item (Bauer 1983, 48). Moon (1998) noted that institutionalization is measured quantitatively in corpus linguistics, and that most institutionalized phrases are rather infrequent (6). *Fixedness* is a complex phenomenon that involves some sort of lexicogrammatical restriction on a lexicalized phrase, one that often limits the scope of the phrase's aspect, mood, or voice (6–7). Finally, *non-compositionality* refers to the inability to determine the institutionally accepted meaning of a lexicalized phrase through a word-by-word analysis of the phrase, and this is typically evident in metaphorical lexicalized phrases (Moon 1998, 7).

Lexicalization of similes, as evidenced in the literature, appears to be a very common occurrence. In his work with prepositional *like* in the British National Corpus (BNC), Hanks found that a “few dozen” simile vehicles recurred frequently. That is a few dozen “out of many thousands of nouns that could in theory function as secondary subjects in similes” (Hanks, 5), to form what he refers to as conventional similes. Additionally, in accordance with the prerequisite of institutionality, Hanks noted that the recurring vehicles form strong associations with specific properties and functions connected

with cultural reference points. He argued that because of this strong cultural connection, many lexicalized similes are considered to be idioms.

Wikberg echoed much of Hanks's view with regard to the ubiquity of lexicalized similes in the BNC. Using the example, "Ruth dropped like a stone into the armchair," Wikberg showed how lexicalized similes can also be formed through the combination of *verb like* constructions and nouns. She provided evidence that this example was a lexicalized simile with 12 instances out of 25 hits in which *stone* followed *drop like a* in the BNC (Wikberg, 128). Numerous other examples of lexicalized similes, in which the adverbial colligates with the verb to describe typical behaviour of certain objects, were also identified in the BNC. These include *collapse like a pack of cards*, *grin like a Cheshire cat*, *sweat like a pig*, *shake/tremble like a leaf*, *smoke like a chimney*, and *sleep like a baby* (Wikberg, 139). As can be seen in the examples provided here, there is seldom any need for elaboration or specific explanation of the grounds with lexicalized similes. By contrast, when a novel simile is produced, elaboration or specific explanation of the grounds is often necessary for the intended meaning of the simile to be conveyed (Wikberg, 130).

3.5.3 Analogical Similes

Another type of simile that is discussed in the literature is the analogical simile, which is a statement of similarity between relations to objects rather than between objects themselves (Ortony 1979, 175). Wikberg (138) found several proverbial analogical similes, which she describes as generic, based on the *is like a* simile pattern in the BNC, one such example being, "An apple pie without cheese, is like a kiss without a squeeze" (AHN [world affairs]). She noted that a common feature of this type of simile involves the

matching of two situations, with both lacking some important or crucial element, and that some are explicitly described as being “sayings.”

3.5.4: Metaphorical Similes

In identifying types of similes, we may return briefly at this point to our discussion of the relationship between simile and metaphor. Whether complicating or adding further credence to the argument that similes and metaphors are distinct linguistic phenomena, it has been found that metaphors are often “mingled inextricably” with similes (Bredin 1998, 75). This leaves us with a category that we must say belongs to metaphorical similes.

One common pattern of metaphorical similes involves a metaphor imbedded as the vehicle of a simile, as seen in the following example from the COCA:

(3.5) By tonight we'll **be like** peas in a pod.
(FIC [FantasySciFi])

The *verb like* element of a simile can also be metaphorical as seen in this analogical example from the BNC, provided by Wikberg (138), in which can also be perceived the conceptual metaphor of stress as liquid:

(3.6) Learning to relax **is like** pulling a plug at the bottom of your stress glass, gradually the tension and stress drain away. (EB1 [applied science])

3.6: Analyzing Similes

Currently, with the notable exceptions of Hanks, Moon (2011b), and Wikberg, there are few studies in the literature that focus specifically on *like* similes, removed from the simile/metaphor discussion. While the research of both Hanks and Wikberg are corpus-based, Moon's study is predominantly a text-analytic study. All three of these papers reveal much about the

behaviour of prepositional *like* in the construction of similes and will be outlined in the following sections, as they play an integral part in the foundation of the current study.

3.6.1: Phraseology

In her comprehensive, corpus-based study, Wikberg examined four major simile structures in the BNC, with the aim of gaining “more specific knowledge than we have had so far of different simile structures, their frequency and communicative functions” (128). Noting that the majority of similes fall into a rather limited range of syntactic patterns, she scanned the BNC for similes showing *as ADJ/ADV as*, *is like a(n)*, *is like V-ing a(n)* and *V like a(n)*.

Regarding the overall degree of figurativeness and innovation, Wikberg (138–139) found notable differences in variation among the four patterns. The most frequent pattern observed in the BNC was the *as ADJ/ADV as* construction. However, the vast majority of these turned out to consist of literal comparisons, while the few instances of this pattern that represented true figurative simile tended to be lexicalized with little innovation. In contrast, the majority (over 90%) of occurrences in the BNC featuring *is like a* were found to be figurative, i.e., true similes, and most of these tended to be innovative (137). Meanwhile, the pattern of *is like V-ing* was the least frequent, but it was often used to form analogical similes.

As noted previously, the *V like a(n)* pattern of adverbial simile consists of predicative and adverbial subcategories. The predicative group, made up of the verbs of perception, was most frequent, accounting for over 50 percent of all occurrences, and these were the only verbs to occur more than once per million words in this pattern. This *V like a(n)* pattern also had

a higher proportion of true similes that were both lexicalized and innovative (Wikberg, 138–139).

The adverbial subcategory of *V like a(n)* was found to occur with over 1,500 different verb forms, and the majority of these only occurred once or twice in the BNC. After sorting through this data, Wikberg (139) found just over 39 percent of this pattern to consist of true simile usage. Many of these verbs were used to describe “events, states and very specific types of behavior” (p. 139), with the vehicle of the simile adding precision to the description. The most commonly used verbs in the *V like a(n)* construction consisted of *behave* and *act*. These verbs need adverbs or prepositional phrases such as expressions with *like*. While they are often used literally, Wikberg found many figurative and lexicalized examples of this construction in the BNC, including *BEHAVE/ACT LIKE a(n) angel, bear, bull, child, monster, pig, and shithead*. The majority of the more than 1,000 verbs in the *V like a(n)* pattern that occurred only once in the BNC were used figuratively. Some examples provided by Wikberg include the following:

- (3.7) The smell of juniper, burned earlier as incense,
still **lingered like** the warm breath of the forest
in summer. (BNC-BNU [world affairs])
- (3.8) Desire **licked like** a flame at her senses, sending
dangerous signals to her brain. (BNC-JXW [imag.])
- (3.9) All this was to seem academic by the Seventies
when the boom finally stopped and the dream
future of abundance and leisure **popped like** a
rosy soap bubble. (BNC-ACS[leisure])

3.6.2: Recurring Verbs and Semantic Classes

In his examination of *like* using the BNC, Hanks identified small groups of recurring verbs and nouns that co-occurred with similes using *like*. As with the findings of Wikberg, the majority of verbs that preceded *like* in similes

were found to be the verbs of perception, *look, sound, taste, and smell*. According to Hanks (1), “*like* has a central role to play in cognition, more powerful than metaphor and going far beyond mere comparison of similarities.” This finding can be seen to fit within the work of Viberg (1983) concerning the unidirectional path for semantic extensions across senses, proceeding downwards from vision. It also fits with and Sweetser’s (1990) findings, where vision is considered the primary sensory modality used for metaphors of knowledge and thought. More recently, Winter (2019, 45) has confirmed this, noting that “vision is ... more efficiently codable, as evidenced by the fact that visual concepts have higher token frequencies and are processed more quickly.” He also notes that taste and smell are relatively ineffable, which refers to the difficulty of verbalizing certain experiences.

In his analysis, Hanks (5–8) identified frequently recurring sets of nouns that collocated with *like a/an N* in the BNC and fell within seven categories of conventional simile. When the target of the simile is a person, three human roles are involved in semantic groups as the most frequent vehicles. The first of these is status (e.g., *child, father, mother*) in relation to the individual, followed by role (e.g., *actor, queen, prisoner*) in society; the next belongs to attributes (e.g., *fool, idiot, drunk*); finally, animals (e.g., *dog, cat, bird*). The wide use of animals as a vehicle for *like* similes was also observed by Wikberg (129), where she notes that “animals represent a much wider variety of creatures and therefore make up a rewarding source of comparison when describing people and their behaviour.” Hanks (5–8) found other frequently occurring semantic groups of nouns, less frequent than the top four listed above, collocating with *like*. These included, in order of frequency, artifacts (e.g., *rocket, light, drum*), events (e.g., *wave, game*,

accident) and irrealis including both events (e.g., *dream*, *nightmare*, *miracle*) and entities (e.g., *ghost*, *angel*, *zombie*).

3.6.3: Patterns of Target–Vehicle Dissimilarity

Moon took a different approach from that of Wikberg and Hanks in her analysis of similes. Rather than looking at phraseology or recurring semantic sets, she examined the target–vehicle dissimilarity patterns of similes in the texts of three 19th-century explorers.

Moon acknowledged that the systematic categorization of targets, vehicles, and grounds of simile (beyond a very general level) would prove to be difficult due to more delicate analyses being “too individualized” (p.140). Instead, she recommended a focus on patterns of dissimilarity through a classification of target–vehicle relationships which, she asserted, may identify points of interest within a text. In her words, “the dissimilarities within similes reveal much, particularly with respect to ideological meanings on the one hand, and the expression of certainty and uncertainty on the other” (p. 134).

Moon provided examples of 12 patterns of target–vehicle contrast, which ranged from physical and concrete to more abstract comparisons. These included the following:

- (1) inanimate = animate
- (2) animate = inanimate
- (3) human = creature
- (4) emotional = physical
- (5) natural environment = built environment
- (6) nature/wild = culture/civilized
- (7) solid = not solid
- (8) living = dead
- (9) beginning = end
- (10) dark = light

(11) real = unreal

(12) good = evil

3.6.4: Functions of *Like* Similes

The main functions of simile usage discussed in the literature centre around explanation and description, including the expression of emotion and illustration making descriptions more vivid and entertaining. When used for general explanation, similes most commonly describe the unknown or unfamiliar in more familiar terms, and this corresponds to what Moon (142) found in her analysis of three 19th-century travel writers.

When used to describe personal feelings, Hanks claimed that *like* similes are “more or less conventional” and “play an important role in presenting the inner and unknowable feelings of an individual to the outside world” (2005, 9). He provided several examples of the *FEEL like* pattern of simile, and he noted that this construction is often used by fiction writers to describe the feelings of their characters.

When the purpose of a simile is for creative description, the conventional format of using the known to describe the unknown does not always apply. Hanks discussed this phenomenon in some detail, providing evidence from the usage of irrealis nouns as vehicles for simile.

The irrealis group of simile makers is further evidence that similes do not invite interpretation [of] one thing in terms of familiar experience, but rather they invite the reader or hearer to categorize one thing in terms of a (sic) salient properties conventionally attributed to another thing within a particular language or culture. (11)

Examples from the BNC provided by Hanks to illustrate this point include the following:

(3.10) The loud, coarse voice ripped through the quiet air with shocking force, **like** a dagger through silk. (BNC-HH1[W_fict_prose])

(3.11) I was so tired I was **like** a zombie.
 (BNC-FU1[W_non_ac_soc_science])

In addition to having unrealistic vehicles, Examples 3.10 and 3.11 are both lexicalized. Hanks argued that while daggers are indeed real objects, it is likely that very few people have direct experience with daggers being used to rip through silk. And, of course, most people would likely agree that zombies are not real, and therefore most people would have no more real experience with them than with daggers ripping through silk. These lexicalized similes are effective because they exist in the collective cultural knowledge of the speech community (Hanks, 11). Hanks made a similar claim about the usage of verbs of perception in *like* similes, noting that they are used to create similes in which the vehicle “invokes an appeal to cultural stereotype rather than to an actual experience of reality” (4). Moon (148) came to a similar conclusion when examining Conrad’s (1899/2007) *Heart of Darkness*, finding that the narrator tended to use simile to describe the familiar in unfamiliar terms.

3.6.5: Simile and Genre

Previous studies have identified several notable differences in how similes are used in different genres. While similes are likely found in most genres, if not all (including prose, poetry, and conversation; Fadaee 2011, 22), previous studies have shown that similes are most commonly found in fiction (Wikberg, 129).

Moon, in her case studies of dissimilarity in text, looked at and compared similes found in travel writing with those found in fiction. The analyses revealed that the main purpose of the similes in travel writing was to describe something unfamiliar in terms of something familiar, while the similes in fiction generally worked in the opposite way, describing the familiar in

unfamiliar ways. She concluded with the cautious supposition that these findings may reflect a wider generic distinction between fiction and nonfiction, but she held back, noting that this was too big of a claim for the scope of her paper.

In terms of patterning and genre, Wikberg (140) observed a difference in usage between the *is like a(n)* and *V like a(n)* constructions. While the proportion of true simile usage was found to be high in the BNC, formed with both *is like a(n)* and *V like a(n)*, notable differences in distribution were observed among the genres. In particular, similes using *is like a(n)* were strongly favoured in the domain of “belief and thought,” and those using *V like a(n)* were strongly favoured in the “imaginative,” “leisure” and “arts” domains. Language related to Wikberg’s category of belief and thought tends to be more abstract, suggestive, and explanatory, and it is therefore more likely to use *is like a(n)* similes. What Wikberg termed imaginative discourse, on the other hand, tends to be used more for description, including evaluation, and the expression of emotions. (Wikberg, 140). Supporting evidence for this explanation can be found in Biber et al. (1999), where the authors demonstrated that two verbs of perception, *look* and *feel*, are used most in fiction while they are used least in academic text.

3.7 Conclusion

The purpose of this chapter has been to provide a comprehensive overview of simile usage with *like*. I started with an operational definition of simile, comprised of a four-point framework collated from previous studies on simile. I then reviewed and summarized the main points of discussion and notable findings from the literature. There is a considerable amount of research that has been carried out on simile, but the vast majority has been concerned with the relationship between simile and metaphor. In many of

these studies, simile is seen as secondary to (or a subset of) metaphor. There is now, however, a growing body of evidence supporting the view that simile and metaphor are distinctively different—most notably, the categorization theory of simile proposed by Glucksberg (2006b). The current research (my thesis) follows this assumption, whereby simile is a separate and distinct phenomenon from metaphor, and because it is underrepresented in the literature on metaphorical language, it is worthy of further and more intensive analysis. An additional gap in the literature, further validating the importance of this line of research, is that only Hanks and Wikberg have as yet presented corpus-driven research, in corpus-based studies, of *like* similes.

In preparing this research for my thesis, it was necessary to review the various types of *like* simile previously identified and discussed in the literature. It was also necessary to review the various methods used by past researchers to analyze and categorize simile. This presented somewhat of a problem in that there has been little agreement among linguists (or in the field of applied linguistics) on the methods and terminology used for analyzing similes and metaphorical language in general.

After identifying and listing the various types of simile that could be expected in the data, I decided to use the combined methodologies of Wikberg, Hanks, and Moon for the present study. Using the grammatical patterning approach to categorize types of simile outlined by Wikberg as the foundation of this analysis, I will examine recurring verbs and semantic sets (Hanks) as well as the patterns of target–vehicle dissimilarity observed in the data. I hope to provide a better understanding of how *like* is used to form similes in real-life language use, as reflected in the corpus data. While Hanks and Wikberg both used the BNC exclusively in their studies, I hope

to be able to provide further insight into the behaviour and usage of *like* similes with comparative data. This will be feasible because the study I conduct for this thesis is to be based on the COCA.

4

Methodology

Corpus-Driven and Corpus-Based

4.1: Introduction

This chapter outlines the methodological orientation and research techniques I developed for this thesis. First, in Section 4.2, I will discuss in detail the two main research questions that I mentioned briefly in Section 1.4. Next, in Section 4.3, I will provide a detailed discussion of the two corpora used in this study: the large reference corpus, the Corpus of Contemporary American English (COCA), and a pedagogic corpus, the English Language Teaching Coursebook Corpus (ELTCC). In Section 4.4, I will discuss the various research techniques used for data analysis in this thesis, including the corpus search functions, concordance sampling techniques, and functions for collocational analysis. Finally, in Section 4.5, I will outline the two main statistical measures that were used in this study.

4.2: Research Questions

As I have mentioned in Section 1.4, this thesis will focus on two main research questions. These are the questions I have posed to investigate the highly frequent word *like* and its usage.

- (1.) What are the most frequent uses of *like*, as reflected in a large reference corpus (COCA)?

(2.) How is *like* treated in English language teaching materials, as reflected in a pedagogic corpus of ELT coursebooks (ELTCC)?

The rationale behind Question 1 is pedagogic in nature, and it is based on the notion of the lexical syllabus first introduced by Sinclair and Renouf (1988, 148). They argued that, for any learner of English, the main focus of study should be on the most common word forms in the language, their central patterns of usage, and the combinations which they typically form. As mentioned in Section 1.1, I chose the word *like* for this case study based directly on Sinclair and Renouf's lexical syllabus, in that *like* is one of the most frequently used words in the English language. Moreover, this high frequency is largely due to *like* having several different functions and senses, with numerous patterns and combinations. Thus, by examining a very large reference corpus (COCA), I aim to objectively identify the most frequent functions, patterns, and phrases of *like* that would be most useful for English language learners.

Next, the rationale behind Question 2 is again based on the lexical syllabus in that, as an English teacher at a Japanese university, I have observed a high level of student error in both the reception and production of *like*. Based on this observation, I set out to determine how coursebooks generally treat this highly frequent word, and to determine whether or not students have been presented with the most frequent patterns of usage and combinations associated with *like*. Following this rationale for Question 2 leads to an important sub-question:

(2.1) Are there notable differences in the treatment of *like* between natural language use as reflected in the COCA, and pedagogic materials as reflected in the ELTCC?

By answering these two main research questions, along with the one sub-question, I hope to raise awareness, among language teachers and

material writers, regarding the importance of incorporating the main principles of the lexical syllabus discussed here.

4.3: Corpora

Two corpora were used for this study, a large reference corpus (COCA) and a pedagogic corpus (ELTCC). The COCA is used to answer the first research question, and the ELTCC is used to answer the second research question, as well as, its corresponding sub-research question.

4.3.1: COCA

The COCA is a very large corpus consisting solely of American English. As of March 2020, it contained more than one billion words in eight different genres. According to its creator, Mark Davies (2020), the COCA is one of the most widely used corpora in the world, and the only corpus that is large, recent, and genre-balanced. While used here as a reference corpus, the COCA is also considered to be a monitor corpus, as it is periodically expanded with new text. Furthermore, it is a balanced corpus with a deliberate sampling frame of data (20 million words) from each year starting in 1990. It currently reflects each year up to 2019—and, because of this, the COCA can be used as a diachronic corpus, to investigate language change over time.

The COCA is currently divided into eight balanced genres as shown in Table 4.1, reproduced from the COCA website (Davies 2020). The COCA has expanded significantly since its initial development in 2008, at which time it consisted of only five genres (spoken, fiction, magazine, newspapers, and academic) and a total of 400 million words (Davies 2010). The most recent update included the addition of three new genres (general-purpose

websites, blogs, and TV/movie subtitles) which put the total size of the corpus over one billion words.

Table 4.1: Genres in the COCA, including number of texts, number of words, and description

Genre	# texts	# words	Description
Spoken	44,803	127,396,932	Transcripts of unscripted conversation from more than 150 different TV and radio programs (examples: <i>All Things Considered</i> [NPR], <i>News-hour</i> [PBS], <i>Good Morning America</i> [ABC], <i>Oprah</i>)
Fiction	25,992	119,505,305	Short stories and plays from literary magazines, children's magazines, popular magazines, first chapters of first edition books, 1990–present, and fan fiction.
Magazine	86,292	127,352,030	Nearly 100 magazines, with a good mix from among specific domains like news, health, home and gardening, women, finance, religion, sports, and so forth
Newspapers	90,243	122,958,016	Newspapers from across US, including <i>USA Today</i> , <i>New York Times</i> , <i>Atlanta Journal Constitution</i> , <i>San Francisco Chronicle</i> , and so forth; good mix from among sections of each newspaper, such as local news, opinion, sports, financial, and so forth
Academic	26,137	120,988,361	More than 200 peer-reviewed journals, covering the full range of academic disciplines, with a good balance among education, social sciences, history, humanities, law, medicine, philosophy/religion, science/technology, and business
Web (General)	88,989	129,899,427	Classified (by Serge Sharoff) into web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages; taken from US portion of the GloWbE corpus
Web (Blog)	98,748	125,496,216	Texts classified by Google as blogs; further classified into web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages; taken from US portion of GloWbE corpus
TV/Movies	23,975	129,293,467	Subtitles from OpenSubtitles.org, and later the TV and Movies corpora; studies have shown that the language from these shows and movies is even more colloquial/core than the data in actual “spoken corpora”
	485,179	1,002,889,754	

In the initial pilot-study phase of my research for this thesis, I considered three reference corpora: the COCA, the Bank of English and the British

National Corpus (BNC). I chose the COCA as the main reference corpus for two reasons. The first and main reason is that the COCA is the only easy-to-access corpus of American English. I considered this to be important because American English is the variety of English taught in the Japanese school system where I currently work; consequently, the majority of ELT coursebooks used in Japanese universities tend to focus on American English. In addition, as a speaker of Canadian English, the variety of English found in the COCA is more familiar to me than the British English of the Bank of English or the BNC.

The second reason for choosing the COCA over the Bank of English or the BNC is due to the fact that the COCA is by far the largest of the three corpora. It is generally agreed by corpus linguists that corpus size is very important and that, in most cases, the larger a corpus is, the more useful it is for identifying linguistic phenomena (e.g., Sinclair 1991, 18; Hunston and Francis 2000, 16; Davies 2015, 12–18). Davies compared data from the one-million-word Brown Corpus (a first-generation corpus) with data from second-generation corpora containing over 100 million words each, such as the BNC and the COCA. By showing that 83 of the top one thousand most frequent adjectives in the 500-million-word COCA occur only five times or less in the Brown Corpus, he demonstrated the importance of corpus size for lexical analysis. The adjectives he presented include very common words, such as *fun*, *medium*, *terrific*, and *affordable*. The contrast consisted in that these same words all occurred over five thousand times in the COCA (13).

Similarly, the importance of corpus size for phraseological research was demonstrated by Sinclair (2005, quoted in Walker 2008, 86), who noted that the frequency of a lexical sequence tends to be one digit lower for each

additional word added to the sequence. This phenomenon is illustrated in Table 4.2, using two phrases containing the word *like* that are examined in this thesis.

Table 4.2: Frequency data for *like* and the extended sequences of the phrases *and just like that*, and *now that's more like it* in the COCA

Lexical sequence	frequency	Lexical sequence	frequency
like	887,453	like	887,453
like that	54,549	like it	18,947
just like that	2,016	more like it	475
and just like that	210	's more like it	90
		that's more like it	75
		now that's more like it	9

The COCA features a web-based interface that is easy to navigate and use, with a wide range of available searches including searches for words, phrases, substrings, lemmas, parts of speech, and synonyms. It uses a proprietary architecture based on Microsoft SQL Server database management, which allows for extremely fast search results.

One possible criticism of the COCA is with regard to the texts that make up the spoken section. Since the developers wanted one-fifth of the original COCA to consist of spoken American English, they decided on using conversation transcripts that were already available in electronic form, due to the nearly impossible task of creating a corpus of that size by recording and transcribing lectures and conversations. The spoken transcripts that were obtained for use in the COCA consist of unscripted conversations from American TV and radio programs, with a bias toward news and political discourse—and this, according to Davies (2020) prompts three questions:

- (1) Do they faithfully represent the actual conversations?
- (2) Is the conversation really unscripted?
- (3) How well does it represent “non-media” varieties of spoken American English?

Regarding the first question, Davies asserts his confidence that the transcripts represent the actual conversations quite accurately, and he provides two examples of transcripts alongside their corresponding audio recordings. These transcripts show that common features of spoken discourse, such as interruptions and false starts are included in the transcripts. Regarding the second question, Davies points out that although there are formulaic/scripted sequences such as “welcome to the program” and “we’ll now go to a commercial break,” his analysis of these transcripts shows about 95 percent of the conversations to be unscripted. Finally, concerning the third question, Davies acknowledges one key difference between natural conversation and the spoken transcripts included in the COCA. This arises from the fact that the speakers knew they were going to be broadcast on TV or radio, and would therefore have likely altered their speech accordingly, for example avoiding profanity and stigmatized language (2020).

Davies goes on to point out, however, that this same issue would also be present in any spoken corpus created by recording live speakers. Davies concludes this discussion by reiterating the fact that searches of the COCA’s spoken section identify numerous examples of phrases commonly found in spoken discourse, such as *like*_{3.2} discourse marker usage (e.g., *so I was like*; 2020).

Despite Davies’ (2020) counterarguments to the criticisms of the spoken section of the COCA listed above, it needs to be acknowledged that these criticisms are in some regard indefensible. There will always be limitations and drawbacks to using TV and radio transcripts for spoken corpora and, despite the claim that they are unscripted, the fact remains that spoken transcripts do not fully represent natural, strictly casual, spoken discourse as it unfolds in non-broadcast settings. While I am fully aware of the

limitations, the spoken section of the COCA is currently the only option for access to a relatively large spoken corpus and will therefore be used in this study. However, any results reported in this study, obtained from the spoken corpus, will need to be assessed with caution.

One issue with using the COCA needed to be addressed during the research phase of this study. It involves the periodic updates to the corpus. The total word count of the COCA has more than doubled since I began my research in 2014. At that time, it consisted of 450 million words in five sections (spoken, fiction, magazine, newspaper, and academic). It was expanded to 520 million words in 2016, and then to 560 million words in 2019. Most recently, in March of 2020, the COCA underwent a major revision with the addition of new search features, and three new sections (general websites, blogs, and TV/movie subtitles) which brought the total word count to over one billion words. Since the majority of the data collection and analysis had taken place before this last revision, I decided to continue working only with the original five sections of the COCA. Also, to address the changes in the size of the corpus, throughout the reporting of results in this thesis I will state the total word count at the time of data collection. In the most recent searches using the COCA for this study (specifically, those used for data collection in Chapter 10), the total word count of the corpus stood at 612 million words, as only the original five sections were included.

4.3.2: The ELTCC

The ELTCC is a specialized corpus of ELT coursebooks, consisting of just over one million words. Since it is made up of coursebooks, it can also be considered a pedagogic corpus, which Willis (2003, 163) defined as a body

of language produced for language learners to process for meaning.¹ The corpus was compiled for the current study, in order to answer the second main research question: How is *like* treated in English language teaching materials, as reflected in a pedagogic corpus of ELT coursebooks?

The ELTCC data consists of 34 complete coursebooks in 13 different series, ranging from single books to four-level courses, all commonly used in Japanese private language schools and universities. Several large international publishers are represented in the ELTCC, and the coursebooks are all intended for general-purpose English, as opposed to English for specific purposes such as business English or academic English. In addition, they all claim to follow communicative language teaching methodology, and they are focused either on speaking or on a more balanced four-skills approach. See Appendix 1 for a full list of the coursebooks in the ELTCC.

Due to time constraints in the research process, only coursebooks that were available online in PDF form were used to compile the data for the ELTCC. I initially sought permission to use the coursebooks but, having received no response from the first few publishers, I decided to continue the data collection under the assumption of fair use (bypassing permission from the copyright holder under certain circumstances). Hilton explains fair use as follows:

If the use of a work furthers progress in the sciences and the arts (i.e. if it promotes learning, knowledge, and the public good) and if its use will do relatively little harm to the author's property rights, then it is not necessary to get the author's permission to use the work. (2001, 50)

1. Willis (1993, 92) used the term *learner corpus* rather than *pedagogic corpus* in an earlier paper. Here the later term *pedagogic corpus* is used.

The coursebook PDFs were first scanned using ABBYY FineReader OCR software, and the scans were then converted to plain text (.txt) files. The corpus was analyzed using two freely available concordance software programs, AntConc (Anthony 2018) and CasualConc (Imao 2018), which both run natively on the macOS computer platform that I use. AntConc, which I used as the primary software for analyzing the ELTCC, has several corpus-analysis tools, including concordances, n-grams, collocates, and word lists. CasualConc is a simpler program with fewer features than AntConc. It was used as a secondary platform, primarily for the collocation feature (see 4.4.3), which displays the data in a visual form, similar to the picture function found in the Bank of English.

The ELTCC is not tagged for parts of speech. Although there is a parts-of-speech tagger available as an add-on for Antconc (TagAnt; Anthony 2015), I was unable to use it with the ELTCC as it repeatedly froze and crashed on my computer. Considering the relatively small size of the ELTCC, this was not considered to be detrimental to the study. I therefore abandoned the tagging process.

4.4: Research Methods and Techniques

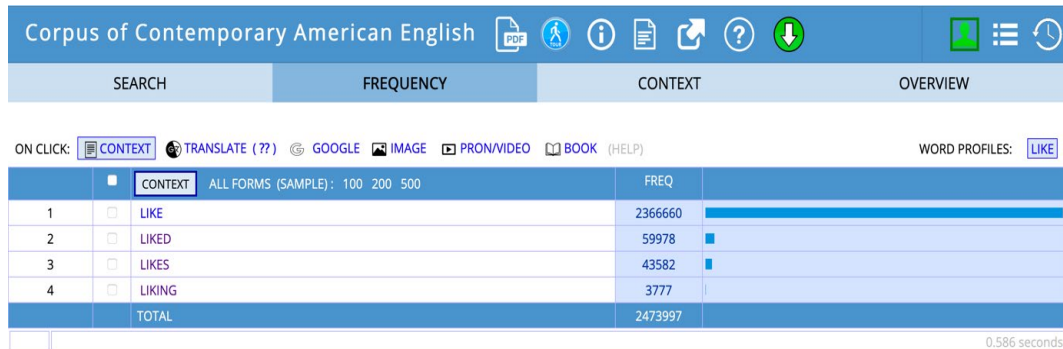
For this study, both the COCA and ELTCC were examined using a hybrid methodology that consisted of both *corpus-driven* and *corpus-based* techniques. Corpus-driven research involves starting with an inductive analysis of the corpus and noting any linguistic features as they emerge. In other words, the corpus data are allowed to speak for themselves, without any preconceived constraints. Corpus-based research, on the other hand, involves looking at pre-selected items of particular interest and analyzing the corpus data to determine how the items behave, generally within a specified linguistic framework. (See Tognini-Bonelli 2001, chaps. 4–5 for further

discussion on these two approaches.) Thus, in this thesis, the corpus-driven phase involved choosing a highly frequent word, *like*, from the COCA frequency list and then examining large concordance samples of *like* to identify the most frequent phrases and patterns of usage associated with it. Once these phrases and patterns had been identified, a corpus-based methodology allowed me to examine in more detail the most common phrases and patterns associated with *like* within linguistic frameworks such as simile usage and phraseological analysis. This methodology is very similar to the commonly-used methodology outlined by Gray and Biber (2015, 127). The next section will outline the various corpus tools and techniques used to carry out the hybrid research methodology I adopted for this study.

4.4.1: List and Chart Displays

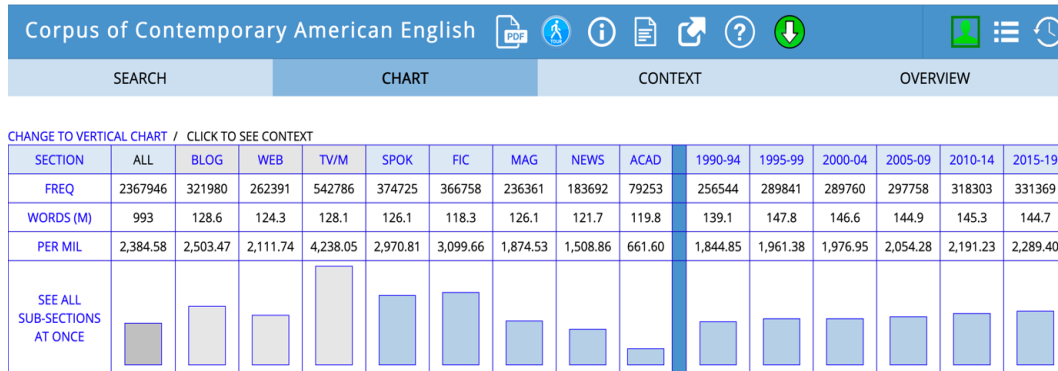
During the initial stage of research using the COCA, I used the list and chart display features to obtain frequency information for the target word and related phrases. The list display simply provides total frequency information, and the user can choose to search the whole corpus or can select which specific sections to search. This feature works for single words, phrases, patterns, and also for lemmas and parts of speech. For example, a search using the term “like” returns the total frequency of the word *like*, while a search using the term “like_v*” returns frequency data for all instances of *like* tagged as a verb. The frequency of phrases such as *that’s more like it* can be obtained easily, and searches focusing on patterns such as *VERB like a NOUN* will provide a list of the searched-for pattern variation, sorted by frequency. Similarly, lemmas can be searched by entering the target lemma in capitals, where a search using the term “LIKE” returns a list of all forms of like (*like, likes, liked, liking*), also sorted by frequency. An example of the COCA list display for the lemma *LIKE* is provided in Figure 4.1.

Figure 4.1: Screenshot of the list display in COCA, showing frequency results for all forms of the lemma *LIKE*



The COCA chart display allows for all the same search functions as the list display, but it outputs the data in a bar graph format. In this way, it shows frequency data for each section of the corpus as well as for the whole corpus. This display also provides frequency data for the search item from each four-year time frame, from 1990 to 2019, allowing the user to investigate how an item's usage has changed over time. In addition to raw frequency, the chart display also provides words-per-million figures, which is necessary for normalizing the data, allowing for comparison with other corpora. Figure 4.2 provides an example of the COCA chart display, showing results for *like* for all eight sections of the newly updated COCA as of March 2020.

Figure 4.2: Screenshot of the COCA chart display, showing results for *like* from each section of the corpus



4.4.2: Concordance Sampling

One of the main research techniques used in the current study, concordance sampling involves obtaining, for the purpose of examining a specified linguistic feature, a randomized sample of concordance lines from a corpus. The most common corpus tool used for concordance sampling is referred to as Keyword in Context and usually abbreviated as KWIC. The keyword (or node) in most corpora can be either a single word or a phrase, and some corpora allow lemma to be used as well (COCA allows lemma searches, but AntConc does not). The node is displayed in the centre and aligned vertically with the selected number of concordance lines running down the screen, and the co-text of the node can be sorted alphabetically, either to the left or the right. This KWIC view allows the researcher to visually scan the concordance lines to identify recurring patterns of usage associated with the node material.

While Sinclair (2003, xv) recommended starting with smaller sample sizes of no more than 100 concordance lines, the default sample size used in the current study consisted of 500 lines. This larger sample size was used in order to make more accurate generalizations of the data with regard to percentage distributions of certain linguistic features. Due to time

constraints, however, smaller 100-concordance-line samples were taken for analyses that required more detailed line-by-line examination for each COCA section (see 5.4). As there are five sections in the COCA, this still consisted of 500 concordance lines, but it was sorted into 100-line samples for access to each section. Figures 4.3 and 4.4 show examples of concordance sampling results from the COCA and ELTCC respectively.

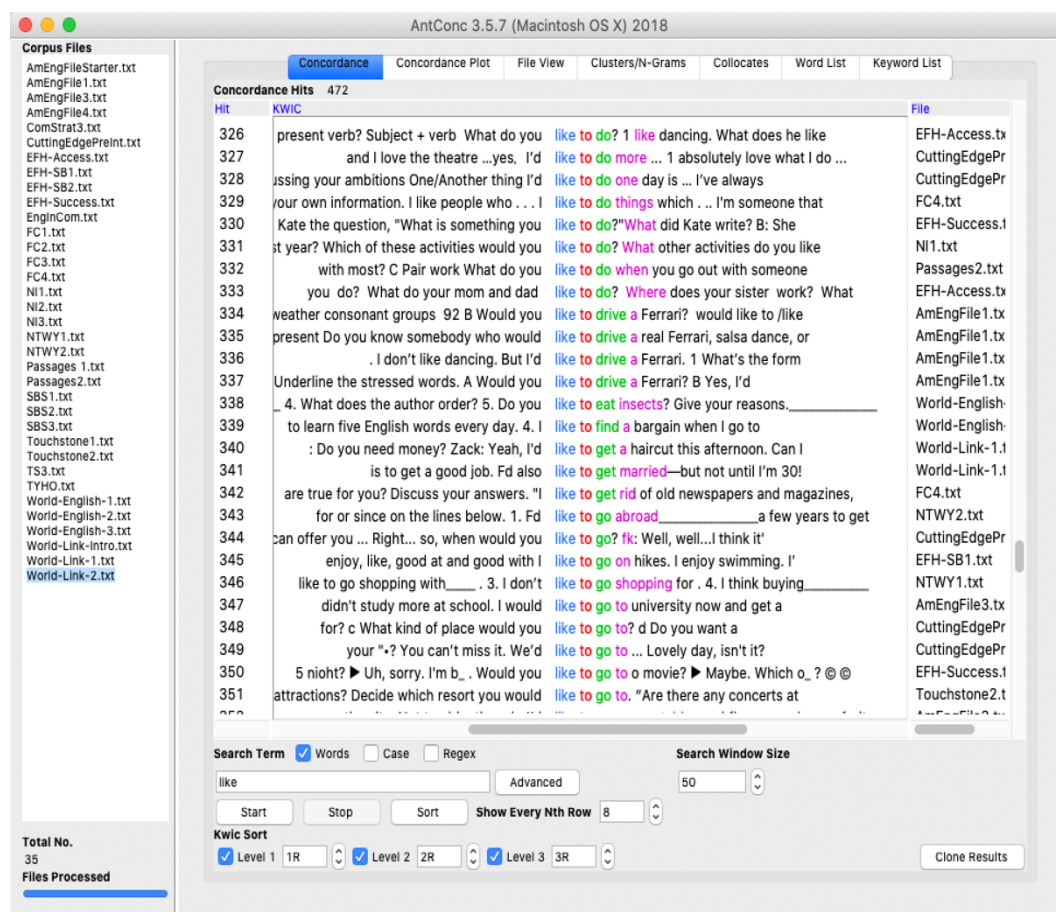
Figure 4.3: Screenshot of the COCA concordance KWIC interface, with *like* as the node

Corpus of Contemporary American English									
SEARCH			FREQUENCY		CONTEXT			OVERVIEW	
314	2012	BLOG	zombietime.com	A B C	, taste the triumph . # Zombie was starting to sound like the comedian in that last comment " ... is just one big				
315	1991	FIC	Bk:LeprechaunBack	A B C	for what felt like the hundredth time ; You sound like a billows , " Silas teased her as he came in from				
316	2019	TV	DC's Legends of Tomorrow	A B C	, that little baby dragon changed my world You sound like my dad 's pitch for ... HeyWorld ! Oh , my God				
317	2003	MOV	Radio	A B C	was riding through there and i heard this noise Sounded like a rabbit or something , you know , caught in a trap				
318	2018	TV	MacGyver	A B C	little girl . But it is hard to shake off sounding like a Yank . When were you in Afghanistan ? Same as you				
319	2012	WEB	...liltalk.nbcsports.com	A B C	# Have n't seen the game yet , but it sounds like Ponder stunk . I wonder how much is defenses figuring out the				
320	1994	MOV	Little Big League	A B C	what I meant . Well , that 's what it sounds like . I just want you to know , I had nothing to				
321	2012	WEB	...og.penelopetrunk.com	A B C	? That does n't sound right to me it sounds like he picked 1982 as cut-off simply because that would make you				
322	2012	WEB	wbur.org	A B C	not allowed in the bar until then ; # it sounds like although you appreciated what this place was , you felt that it				
323	2012	BLOG	metafilter.com	A B C	with you i 'm just going to restate something that sounds like what you said , but dumb and clumsy on purpose . i				
324	2014	MAG	PopMech	A B C	the anchor in the wall . # ART PESCHKE # SOUNDS LIKE YOU 'RE REALLY CHOPPING HAIRS HERE # The author of " The				
325	1995	TV	Star Trek: Deep Spac...	A B C	a writer . I 'm not a writer yet it Sounds like you 're waiting for something to happen that 's going to turn				
326	2012	BLOG	permaculturenews.org	A B C	food security particularly where low-protein starch staples like cassava and corn dominate the diet . # Chayote (Sechium edul				
327	2008	FIC	Bk:ShadowReichenbach	A B C	falls were gossamer at the top , then line and striated like a mare 's tail , and then massive and columnar as they				
328	2007	SPOK	CNN_King	A B C	course , it could happen in Des Moines . And stuff like that has happened . I mean , listen , some of the				
329	2012	WEB	goodreads.com	A B C	for a long time and was able to take a subject like this and transform it into a a good read . He does				
330	2012	BLOG	shoemoney.com	A B C	your success , i ca n't imagine going through surgery like that . Much respect Shoemoney . May you have continued suc				
331	2012	BLOG	almanacofeats.com	A B C	is a great candy bar , but it does n't taste like a Milky Way , " my mom commented . " It actually				
332	2012	BLOG	farine-mc.com	A B C	... outside of sausage . It was so tasty but tastes like the real thing . Hope this is what you looked for .				
333	2002	FIC	SouthwestRev	A B C	putting something in my mouth under my tongue it tastes like dirt , but i suck at it as though it were as				
334	2012	WEB	nightmaremode.net	A B C	Conversation from Facebook Trackbacks # ... got a tattoo like Mass Effects ' JackStuffFive gaming achievements to be				
335	2004	FIC	NewEnglandRev	A B C	him in a snap-brim Borsalino and grinning a set of teeth like perfectly even shells were as much of a staple as the postcards				
336	1997	NEWS	Denver	A B C	's inability to rid the town of junk cars and the like , and complaints that residents were n't allowed to speak at				
337	2012	WEB	...iewsfrothecouch.com	A B C	told that shit from the teacher and i looked at them like they where brain and said really ? so if you do nt				
338	2019	MAG	The Atlantic	A B C	People who donate say that they primarily do so because they like helping others . It feels more direct than just donating to a				
339	2012	BLOG	...ng.blogs.nytimes.com	A B C	just vote from their characteristics , like what things they like , the color of their skin , and how much their worth				
340	2012	BLOG	mmaweekly.com	A B C	they love the sport , they love the rush they like to fight and go down fighting . Title or not , those				
341	2012	BLOG	ianwelsh.net	A B C	a huge problem . They 're trained in violence they like it and they want to keep doing it . If you fire				

Figure 4.3 provides an example of the COCA KWIC concordance interface from a 500-concordance-line search, setting *like* as the node and sorted alphabetically to the left. As you can see, the COCA uses colour coding for parts of speech, and this makes visual analysis of the data easier. On the left of the screen, you can see the sample line number, the year of publication, the COCA section, and the name of the text from which the line

was taken. In this example, we can clearly see the pattern *SOUND like N*, which is highly frequent in the COCA.

Figure 4.4: Screenshot of the AntConc concordance (KWIC) interface, running the ELTCC with *like* as the node



Next, in Figure 4.4, we can see a similar AntConc KWIC search of *like* in the ELTCC, this one sorted to the right of the node, clearly showing the verbal *like to infinitive* construction. The AntConc software is simpler, and it provides less information than the COCA. While it is colour-coded, the coding is not for parts of speech; it rather signifies the sorting level. The node word is shown in blue, the first word to the right is shown in red, the second word is shown in green, and so forth. Also, while the name of the file is provided for each concordance line, additional information such as

year of publication is not shown. Finally, while the COCA allows users to choose the number of concordance lines to input, and while it automatically randomizes the results, AntConc will output all available instances. In order to reduce the output to a manageable number, AntConc allows you to manually thin the results by entering a number in the “Show Every Nth Row” input box.

In order to manage the large amounts of data collected for this study, I copied and pasted all KWIC concordance results into Microsoft Excel files.¹ For certain searches, such as variations of semi-fixed phrases or patterns (e.g., *LOOK like a million bucks/dollars*), it was necessary to combine several separate searches into one Excel file. This was also necessary when combining different forms of a lemma in the ELTCC, as it does not have the capability to include lemmas² in KWIC concordance searches.

After transferring the target data to Excel spreadsheets, the research generally moved to a more corpus-based approach where each concordance line was examined and either assigned a code based on a specific linguistic feature. For one example, I applied the reversibility test to determine whether a *BE like* comparison was a true simile or a literal comparison. If it was found to be tagged incorrectly or to not belong among any of the features under examination, it was removed from the data. For another example, verbal and prepositional *like* share similar patterning with regard to premodification and, due to tagging errors, they are often included in the same search results. So, there were often instances of similar patterning

1. The COCA allows users to save KWIC concordance lists, but this is limited to 100 lines at a time. AntConc does not have a data-saving feature.

2. Several researchers (e.g., Sinclair 1991; Stubbs 1996), due to different forms of a lemma often displaying unique phraseological behaviour, caution against using lemmas in phraseological analysis. However, prepositional verb patterns with *like* were found to show no differences among the various forms of the verb (see Chapter 6).

that nevertheless present different senses of *like*, making it necessary to carry out time-consuming manual sorting and coding of the data. Once this sorting process was completed, the coding column could be resorted in numerical order, which allowed for a tallying of each feature to determine frequency distributions.

4.4.3: Collocational Analysis

Most corpus-analysis platforms, including the COCA as well as AntConc and CasualConc software, feature collocation analysis tools which are useful for examining the phraseological behaviour of target words or phrases. Hunston (2010, 163) noted that collocational analysis can be used as a powerful complement to concordance sampling, and it can act as a shortcut to discovering patterns and phrases that other methods (such as n-grams) may fail to identify. One of the main benefits of collocational analysis is the ability to identify collocations of the target word that may not have appeared in the corresponding concordance samples. Considering that there are 1,240,789 occurrences of *like* in the 612-million-word COCA, it is highly unlikely that all relevant collocations would appear in a sample of 500 lines. Another benefit of the collocation views using the COCA and AntConc software is their inclusion of objective statistical calculations along with the raw frequency data. While the collocation tool included in the CasualConc software was used to help with data visualization, it is only based on frequency and not statistical analysis. (See 4.4.3 for a discussion of the statistical calculations used in the corpus software.)

The collocation tool in the COCA was used to complement the concordance sampling for all areas of data collection in the first part of this study. In my research for Chapter 6, however (which focuses on the *perception verb + like* construction), it was used as the primary source of data collection.

In this part of the study, the collocation tool was particularly useful, as it can identify collocates of words or phrases using statistical analysis. Moreover, it allows the researcher to specify the span (defined by Sinclair [2003, 179] as “the distance between two collocating words”) on either side of the node word. Figure 4.5 shows a screenshot of the results for a concordance search of the *look like* construction.

Figure 4.5: Screenshot of concordance results for the *look like* construction in COCA, sorted by frequency

	CONTEXT	FREQ	ALL	%	Mi	
1	<input type="checkbox"/> GIANT	277	28141	0.98	4.28	
2	<input type="checkbox"/> CROSS	151	34178	0.44	3.13	
3	<input type="checkbox"/> IDIOT	127	4947	2.57	5.67	
4	<input type="checkbox"/> FOOL	116	10268	1.13	4.48	
5	<input type="checkbox"/> TYPICAL	116	28321	0.41	3.02	
6	<input type="checkbox"/> SHIT	110	16845	0.65	3.69	
7	<input type="checkbox"/> ORDINARY	107	20509	0.52	3.37	
8	<input type="checkbox"/> BUNCH	105	18881	0.56	3.46	
9	<input type="checkbox"/> ACCIDENT	99	23734	0.42	3.04	
10	<input type="checkbox"/> MINIATURE	89	5240	1.70	5.07	
11	<input type="checkbox"/> ANGEL	88	11572	0.76	3.91	
12	<input type="checkbox"/> BELONGED	69	8565	0.81	3.99	
13	<input type="checkbox"/> GHOST	66	10426	0.63	3.65	
14	<input type="checkbox"/> TOY	57	9691	0.59	3.54	
15	<input type="checkbox"/> BELONGS	56	7057	0.79	3.97	
16	<input type="checkbox"/> BUCKS	55	10015	0.55	3.44	
17	<input type="checkbox"/> CLOWN	50	3036	1.65	5.03	
18	<input type="checkbox"/> DOLL	50	6630	0.75	3.90	
19	<input type="checkbox"/> DUCK	50	8270	0.60	3.58	
20	<input type="checkbox"/> BELONG	49	11560	0.42	3.07	
21	<input type="checkbox"/> PRINCESS	47	10189	0.46	3.19	
22	<input type="checkbox"/> FOOLS	42	2446	1.72	5.09	
23	<input type="checkbox"/> RAT	42	6791	0.62	3.61	
24	<input type="checkbox"/> PIG	39	7022	0.56	3.46	
25	<input type="checkbox"/> STATUE	39	7844	0.50	3.30	
26	<input type="checkbox"/> GENIUS	39	9388	0.42	3.04	

The results in Figure 4.5 above show the most frequent and statistically significant collocates of *look like*. Since I was specifically searching for the *LOOK like a/an NOUN* construction, I selected only collocates to the right of the node and with a span of four words. This four-word span allows for the detection of non-sequential patterns, such as collocates of *look like* that are premodified with adjectives. The default span in the COCA is four words to the left and right of the node; however, this can be expanded as

desired. While this default was deemed more than adequate for these particular searches, studies by Sinclair and Jones (1974) and Mason (1997, 1999) have found a four-word span to be the optimum for collocation analysis.

The collocation tool in AntConc was used in the second stage of the research to examine the behaviour of *like* in the ELTCC. Like the COCA collocation tool, AntConc allows for collocate searches of both words and phrases, and the user can specify the span and whether to sort the results by frequency or statistical strength. Here, however, the default span is five words to the left and right, and this was reduced to four words as I sought consistency with the COCA searches. An example of AntConc's collocation output is provided in Figure 4.6.

Figure 4.6: Screenshot of AntConc concordance results for *like*, sorted by frequency

AntConc 3.5.7 (Macintosh OS X) 2018

Corpus Files

- AmEngFileStarter.txt
- AmEngFile1.txt
- AmEngFile3.txt
- AmEngFile4.txt
- ComStrat3.txt
- CuttingEdgePreInt.txt
- EFH-Access.txt
- EFH-SB1.txt
- EFH-SB2.txt
- EFH-Success.txt
- EngInCom.txt
- FC1.txt
- FC2.txt
- FC3.txt
- FC4.txt
- NI1.txt
- NI2.txt
- NI3.txt
- NTWY1.txt
- NTWY2.txt
- Passages 1.txt
- Passages2.txt
- SBS1.txt
- SBS2.txt
- SBS3.txt
- Touchstone1.txt
- Touchstone2.txt
- TS3.txt
- TYHO.txt
- World-English-1.txt
- World-English-2.txt
- World-English-3.txt
- World-Link-Intro.txt
- World-Link-1.txt
- World-Link-2.txt

Total No. 35
Files Processed

Concordance Concordance Plot File View Clusters/N-Grams **Collocates** Word List Keyword List

Total No. of Collocate Types: 3446 Total No. of Collocate Tokens: 30188

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	1586	1254	332	4.24643	you
2	1386	179	1207	3.95121	to
3	1263	878	385	4.19509	i
4	828	522	306	4.63098	do
5	806	304	502	2.65441	a
6	698	259	439	2.18054	the
7	602	387	215	4.31169	what
8	580	538	42	6.28654	would
9	490	417	73	4.07040	t
10	477	189	288	2.86453	and
11	459	162	297	3.72242	it
12	323	262	61	5.20364	don
13	303	242	61	5.23159	d
14	293	148	145	2.79658	s
15	289	107	182	3.13637	b
16	278	126	152	2.94349	your
17	276	138	138	4.31769	like
18	256	161	95	2.52297	of
19	256	65	191	2.17517	in
20	201	86	115	3.22788	that
21	198	104	94	2.31792	is
22	177	144	33	4.98785	really
23	171	60	111	3.09765	or
24	170	58	112	2.88849	about
25	164	102	62	3.31193	people

Search Term ☒ Words ☐ Case ☐ Regex

like Advanced

Window Span ☐ Same From... 4L To... 4R

Start Stop Sort

Min. Collocate Frequency 1

Sort by ☐ Invert Order Sort by Freq

Clone Results

The results in Figure 4.6 show how the AntConc collocation tool can be used to easily identify the phraseological behaviour of *like* in the ELTCC. Here, the predominance of verbal *like* over prepositional *like* is apparent. For example, we can see that *to* occurs to the right of *like* 1,207 times, and we can see that *would* occurs to the left of *like* 538 times. Clicking on the collocates brings up all of these co-occurrences in KWIC view, where it can be confirmed that these collocations are components of the *like to* + *INFINITIVE* and *would like* (*like*_{2.2}) verbal constuctions.

Primarily for the visual layout of the collocates in relation to the node, the collocation tool in CasualConc was used alongside AntConc. This is similar to the picture function found in the Bank of English, and it is a feature that the COCA unfortunately does not have. CascualConc is a much simpler platform than AntConc, however, and it does not include statistical analysis. The results are based on frequency information only. Figure 4.7 provides an example of this CasualConc collocate output.

Figure 4.7: Screenshot of the CasualConc collocation visual display for *like*. The red figures represent the frequency of the context word in the position found relative to the keyword.

	Context Word	LR Total	L Total	R Total	L5	L4	L3	L2	L1	Keywo...	R1	R2	R3	R4	R5
1	you	1,498	1,241	257	71	60	79	158	873	0	67	36	49	55	50
2	to	1,372	168	1,204	40	38	42	32	16	0	979	51	68	40	66
3	i	1,078	870	208	41	79	169	233	348	0	32	36	53	53	34
4	do	752	521	231	36	24	57	403	1	0	9	154	20	32	16
5	a	721	328	393	77	100	92	54	5	0	132	31	101	73	56
6	the	676	264	412	70	76	59	58	1	0	162	46	79	65	60
7	would	576	541	35	8	6	30	349	148	0	1	8	10	6	10
8	what	541	426	115	63	170	180	12	1	0	46	19	21	9	20
9	t	469	399	70	14	13	17	71	284	0	2	9	19	20	20
10	and	466	209	257	53	42	53	53	8	0	31	66	72	50	38
11	it	415	147	268	19	17	27	55	29	0	149	33	42	20	24
12	don	299	250	49	2	6	47	195	0	0	4	14	11	10	10
13	d	285	245	40	16	9	4	17	199	0	3	8	13	11	5
14	in	268	78	190	35	19	14	9	1	0	25	27	88	29	21
15	s	258	143	115	23	26	43	39	12	0	1	40	25	25	24
16	of	249	169	80	39	81	29	16	4	0	0	9	20	31	20
17	your	241	114	127	14	16	23	58	3	0	27	21	26	35	18
18	like	234	117	117	39	45	22	9	2	3,734	2	9	22	45	39
19	that	210	92	118	14	18	22	38	0	0	68	13	17	9	11
20	i	185	151	34	24	27	40	35	25	0	5	7	13	6	3
21	or	180	61	119	11	13	13	19	5	0	13	32	31	23	20
22	b	162	129	33	34	28	43	20	4	0	6	8	7	4	8
23	people	160	101	59	5	12	10	20	54	0	10	14	9	14	12

Figure 4.7 shows how the visual layout of the CasualConc concordance tool makes it easy to identify the exact position of the collocates in relation to the node word, with the collocate frequency highlighted in red. This provides useful information when used alongside the AntConc collocate tool, which does not provide the specific location of the collocate in the span.

4.4.3.1: Mutual Information for Measuring Collocational Strength

It is generally agreed (Greenbaum 1974; Stubbs 2001; Hunston 2003) that intuition and frequency counts alone are insufficient for identifying collocation, and most corpus-analysis software includes some form of statistical test, or association measure, used to quantify collocation strength. There are numerous statistical tests used to measure collocational strength: Gries (2015, 56) claimed that more than 80 different association measures have been discussed. However, the most commonly used and discussed association measures tend to be the t-score and mutual information (MI) (Gablasova et al. 2017, 161). Both statistical measures have their own strengths and weaknesses, and because of this, some corpora allow the user to specify which test to use when analyzing collocation strength.

The t-score, measures “how certain we can be that the collocation is the result of more than the vagaries of a particular corpus” (Hunston 2002, 72). It was originally derived from the t-test and is calculated by subtracting the frequency of random co-occurrence from an adjusted value of collocation frequency. It is directly dependent on corpus size and is not standardized, and therefore, it cannot be compared across corpora (Gablasova et al 2017, 162). Because it highlights frequent word combinations t-score rankings are closely related to rankings based on raw frequency (Durrant & Schmitt 2009). The MI score, on the other hand, measures “how strongly

two words seem to associate in a corpus” (Hunston 2002, 72). It is calculated with a logarithmic scale that provides the ratio between the frequency of the collocation and the frequency of random co-occurrence of the items in the collocation (Church & Hanks 1990, 23). Unlike the t-score, MI is normalized and therefore comparable across corpora, with scores of 3.0. or above considered to show a significant collocation (Hunston 2002, 71). Despite MI being the only association measure available in both the COCA and AntConc collocation tools, it was preferred over t-score for its ability to be compared across corpora. According to the developer of the COCA, MI was chosen as the sole measure of collocational strength for reasons of simplicity, as too many measures are sometimes confusing for users. The main criticism of MI is the prominence it gives to words with very low frequency (e.g., 1–3 tokens), but this can be addressed in the COCA through the ability to specify a frequency threshold for the results (Davies, personal communication, 2015). This feature was used in the current study by setting the frequency threshold for collocations to ten occurrences.

4.5: Summary

In this chapter, I have outlined the research methods and techniques that I used to carry out the research for this thesis. I have discussed the rationale behind the two main research questions, and I have provided a detailed overview of the two corpora used to answer the research questions. I have also described the various research techniques that I used to analyze the data and, finally, I have discussed the statistical measures that were used. While this chapter is meant to provide an overview of the general methodology and research techniques used in this study, more detailed information is at times provided in the relevant results chapters. For example, I will go

into detail concerning such things as corpus search items and specific linguistic criteria that were analyzed.

5

Results

Similes with *BE Like*

5.1: *Like* as Vehicle for Simile

In this chapter, through an examination of the *BE like* construction, and a comparison of its usage across genres as reflected in the five sections of the Corpus of Contemporary American English (COCA), I will explore the role of *like* as a vehicle for simile. Two main data sets were used for this study: 500-concordance-line samples of *BE like*, one from each section of the COCA, followed by a narrowed-down examination consisting of 100-concordance-line samples of *NOUN BE like NOUN*, one from each COCA section.

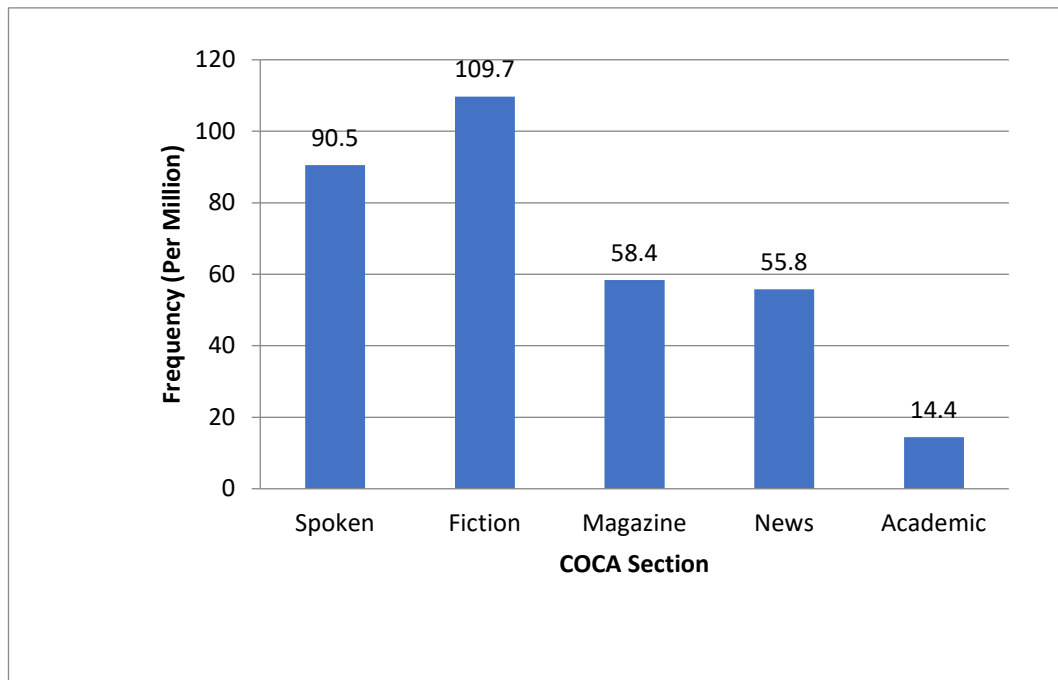
This results chapter is organized into four main sections, as follows. In section 5.2, I look at the frequency data, in each of the five sections of the COCA, for simile usage with the *BE like* construction. I will show how the frequency of simile usage varies widely across genres. The remaining three sections will each focus on a different analysis of *BE like* simile usage in the COCA. In section 5.3, I examine the various types of *BE like* similes identified in the data, and I show how different types of simile are favoured in different genres. In section 5.4, I look at trends in the target–vehicle dissimilarity patterns of *BE like* similes, and I demonstrate that the majority of similes can be categorized into a small number of very common patterns.

However, as I will point out, there are clear differences among genres in how they favour or disfavour specific patterns. Finally, in section 5.5, I explore the rationale and motivation behind *BE like* simile usage across the different genres, and I discuss how these findings compare and relate to Moon's (2011b) research, and to her discussion of simile and dissimilarity in fiction and nonfiction.

5.2: Frequency of *BE Like* Similes in the COCA

As outlined in Sections 1.4, and 3.2.2 *like* is used for several functions besides simile usage such as literal comparison. In 500-concordance-line samples from each COCA section, 52.8% of the data was simile usage in fiction, while only 25.8% was simile usage in spoken, and this lower proportion was due to frequent pragmatic *like*₃ usage. The magazine, news, and academic sections had nearly identical proportions at 39.8%, 40.6%, and 40.4% respectively. Figure 5.1 shows the adjusted overall frequency of *BE like* simile usage found in each of the five sections of the COCA, calculated by multiplying the overall frequencies of *BE like* by the proportions of simile usage identified in 500-concordance-line samples.

Figure 5.1: Overall Frequency of similes by section of the Corpus of Contemporary American English (COCA), from 500-line samples of *BE like*



These data show us that the *BE like* similes are most commonly in fiction and spoken English, and least common in academic English with magazine and news falling in between. These frequency results are not surprising. Simile is an important literary device used by authors to make simple descriptions more vivid and memorable for the reader. It is covered in detail in many writing guides. For example, two such guides are characterized in their titles: *Literary Devices: “All Writers Should Know!”* (Kisak 2015) and *Rhetorical Devices: A Handbook and Activities for Student Writers* (McGuigan, 2007).

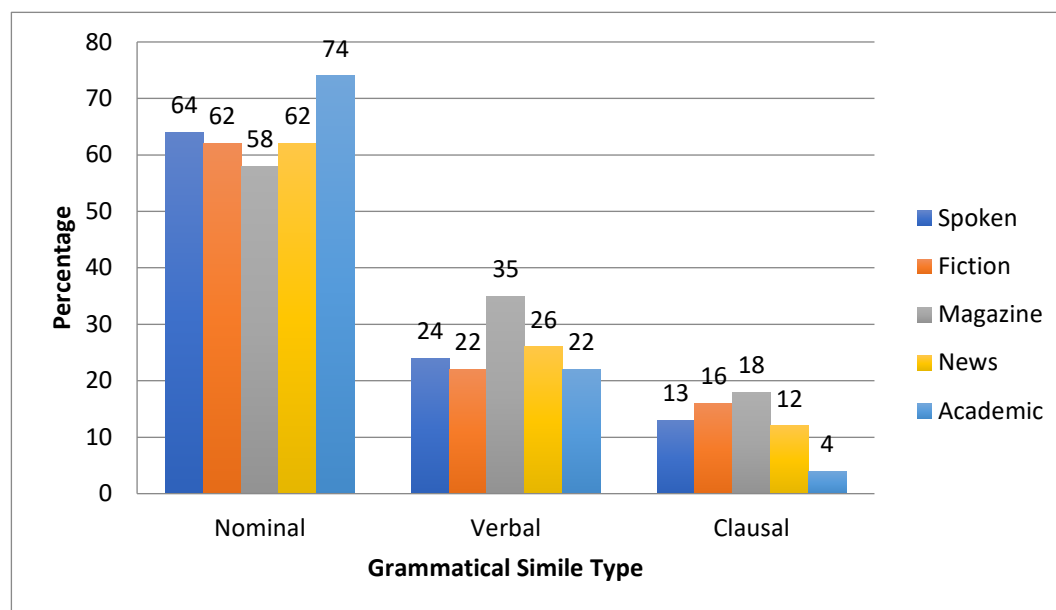
5.3: *BE Like* Simile Type Usage Across Genre

5.3.1: *BE Like* Frequency and Grammatical Structure

As outlined in Chapter 2, similes using *like* fall into three main grammatical categories: adverbial, predicative (including nominal and verbal), and

clausal. But since my focus for now is on similes using the *BE like* construction, in this analysis I will cover only nominal predicative, verbal predicative, and clausal similes. The adverbial *VERB like* construction will be my focus in chapter 6.

Figure 5.2: Percentage frequency distribution for the three grammatical types of *BE like* simile, identified in 500-line samples taken from each of the five sections of the COCA



Looking at the frequency distribution of the three grammatical types of *BE like* similes in Figure 5.2, we can see that nominal similes are twice as frequent as verbal similes. The latter, in turn, are twice as frequent as clausal similes. This tells us that predicative *BE like* similes are used much more often to describe things than to describe actions or situations.

Nominal predicative similes notably more frequent in the academic section. Verbal predicative similes are notably more frequent in the magazine section, and clausal are notably less frequent in the academic section. In the following subsections, alongside detailed analyses of each of these grammatical simile types, I will unpack and discuss these observations.

5.3.1.1: Predicative Similes

Both nominal predicative and verbal predicative *BE like* similes showed up in the data. Nominal predicative similes are compound predicates that consist of a noun or noun phrase following *BE like* (Example 5.1),¹ and verbal predicative similes are simple predicates that consist of *BE like* followed by a gerund or gerund phrase (Example 5.2).

(5.1) My brain **is like** a bowl of scrambled eggs today.

(FIC[Bk:AllGlitters])

(5.2) This must **be like** being hugged by a boa constrictor, Charley added.

(FIC[BkJuv:VoiceWind])

Nominal predicative similes. The nominal predicative type of simile represents the prototypical *A is like B* simile, where one thing is described by comparing it to a different thing. Examples 5.3–5.7, from each of the five sections of the COCA, highlight typical usage of nominal similes across these genres.

(5.3) To me, music **is like** soup. Music comes down to your throat. It feels so warm. (SPOK[CBS_Sixty])

(5.4) Above, the clouds **were like** little white bubbles of soap that had been incompletely sponged off the hard slate of the blue morning sky.

(FIC[VirginiaQRev])

(5.5) The Big Mountain **is like** a giant mixing bowl flipped over, its balding summit quickly giving way to evergreen groves and glades, and somewhere in all of that there's always a stash to find.

(MAG[Skiing])

(5.6) Electricity **is like** cash in the desert. Everybody wants it. Everybody needs it. And I've found myself willing to trade in all I have to get it.

(NEWS[AssocPress])

1. Numbered examples will be designated in the text with the corresponding number in parentheses.

(5.7) Globalization **is like** an economic freight train.
 If one is prepared to jump on board, then one can
 go far—if you pay the price for the ticket.
 (ACAD[ArabStudies])

While it can be argued that all the above examples demonstrate the common simile usage of describing the unknown using the known, Example 5.3 seems to demonstrate another usage, that of expressing an opinion. One's opinion can be considered an unknown that needs to be explained, but it does represent a pragmatic usage of simile. Furthermore, Examples 5.4 and 5.5, from the fiction and magazine sections respectively, illustrate how similes used in creative writing are often more verbose than similes used in other genres.

While nominal predicative *BE like* similes are the most common grammatical simile type across all sections of the COCA, in academic English they turned out to be, on average, twelve percent more frequent than in the other sections. Here, they accounted for 74 percent of the sample. This suggests that academic English is perhaps more predisposed to using objects, as opposed to actions or situations, as vehicles for similes. Perhaps writers consider these nominal constructions to be more easily understood or imagined by the reader than verbal constructions. This seems to make sense, as one would think that the main purpose or goal of any academic text should be to transmit information to the reader. That is, academics need to be especially conscious of the Gricean maxim of manner, whereby obscurity and ambiguity are to be avoided.

Verbal predicative similes. Verbal predicative similes generally function in the same way as nominal predicative similes: to describe something unknown or unfamiliar, in known or familiar terms. The main difference is simply reflected in the grammatical structure, in that a situation or action, rather than a thing, is being used as the vehicle for the simile. The following

examples, from each of the five sections of the COCA, illustrate this simile type.

- (5.8) and that's when Speaker Boehner said it's **like**
negotiating with jello down there.
(SPOK[CBS_FaceNat])
- (5.9) The job **was like** forcing blooms back in to buds.
(FIC[Triquarterly])
- (5.10) Treating them as puzzles **is like** trying to
solve the unsolvable—an impossible challenge.
(MAG[Smithsonian])
- (5.11) It **was like** reading a novel with all the vowels
removed, or watching a film run backward. (News)
- (5.12) Talking with Ned is like falling down the rabbit
hole into a Leibnizian universe.
(ACAD[AmerScholar])

Verbal predicative similes are much less frequent than their nominal counterparts. This type of simile only accounts for approximately one-quarter of simile usage across all sections of the COCA with the exception of the magazine section—which has a notably higher proportion than the other sections at 35 percent. This is possibly due to magazine writers wishing to have a more dynamic, experiential effect with their descriptions.

Clausal similes. In clausal predicative similes, *BE like* is followed by a full clause and can be glossed with *as if* or *as though*. This type of simile is the least frequent grammatical simile type found in the COCA, with only 54 occurrences in the entire sample of 2,500 *BE like* concordance lines. Clausal similes are most commonly found in the spoken and fiction sections; they are notably lower in frequency in the academic section. This is not surprising, as using *like* in place of *as if* or *as though* is often considered improper usage in prescriptive grammars and thus it is discouraged, despite the fact that it has been used in this way since the 14th century (M-W). The fact that it is often frowned upon by prescriptivists may be why it is used

very infrequently in the more formal register of academic writing as it seeks acceptance with a serious audience. The casual, colloquial nature of clausal *BE like* similes is further exemplified by the fact that nearly all occurrences of this structure in the data, across all five sections of the COCA, were attributed to reported or quoted speech (see Examples 5.13–5.17), including all eight occurrences in the academic section.

- (5.13) It **was like** he stuck a knife into my heart.
(SPOK[ABC_20/20])
- (5.14) It **was like** my mind was turning the pages into a movie. (FIC[NewEnglandRev])
- (5.15) It's **like** life holds a secret, and having a child is that secret. (MAG[Redbook])
- (5.16) Nikole felt a weird tightness. It **was like** a band was wrapped around her midsection.
(NEWS[Atlanta])
- (5.17) "I call perfectly managed forests sterile. It's **like** someone came through with a vacuum cleaner over the forest floor." (ACAD[AmerScholar])

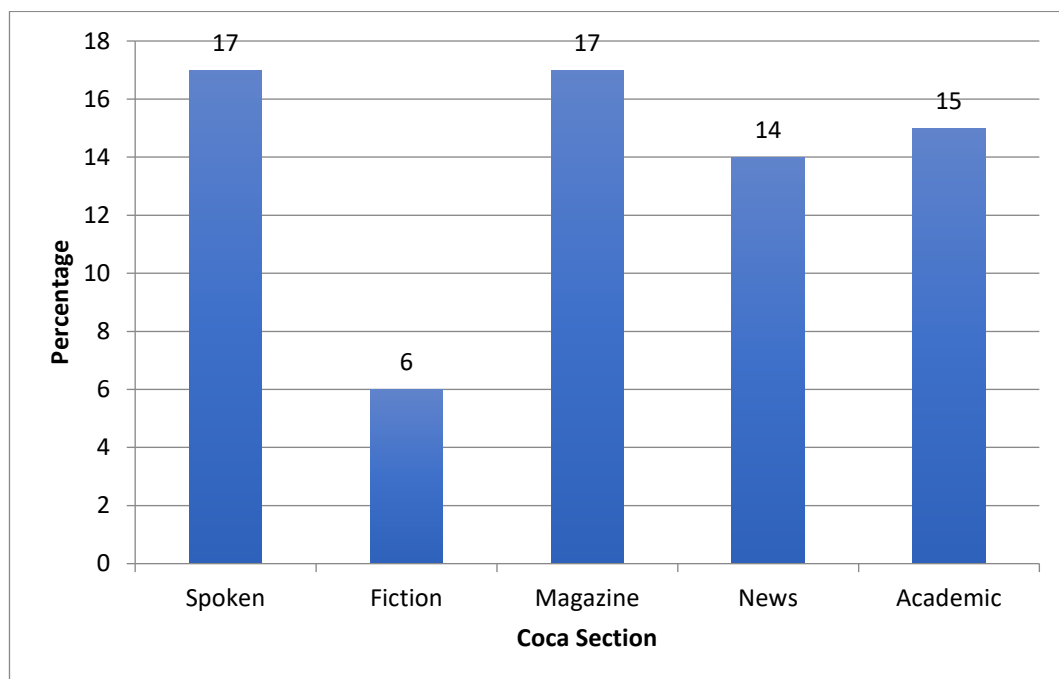
5.3.2: Analogous Similes

An analogous simile is a statement of similarity between relations to objects rather than objects themselves (Ortony, 1979: 175). Many verbal predicate similes are analogical (see Example 5.18), especially when the verb is repeated in both the target and vehicle (Example 5.19). However, nominal predicate similes can also be analogous, as seen in Example 5.20 below.

- (5.18) Shopping on AOL **is like** living in a gated community, says America Online VP Wendy Brown. It's extremely safe and secure. (MAG[Newsweek])
- (5.19) Being in her presence **was like** being on speed. (FIC[KenyonRev])
- (5.20) . . . but the difference between this place and where he lived in Brooklyn **was like** the difference between a wool suit on a hanger, and a lamb. (FIC[Commentary])

Figure 5.3 shows that analogical similes do not occur frequently in the data. Four of the five sections—spoken, magazine, news, and academic—have very similar levels of this type of simile, with an average of 16 percent. Moreover, the fiction section had a lower proportion, with only six percent of its similes being analogical. It is interesting to note that fiction is the most simile-heavy genre, yet the least likely of all the genres to use analogous simile. Perhaps, this type of simile is considered too basic and structured for a more creative writing style.

Figure 5.3: Percentage of analogical similes by section of the COCA



5.3.3: Lexicalized Similes

As I will outline in Chapter 6, the data suggest that *VERB like* lexicalized similes are as much as four times more common than their *BE like* counterparts. However, a number of lexicalized *BE like* similes were identified in the data, using the collocates search feature of the COCA, and these too were found to show differences in usage across sections of the COCA.

Table 5.1 shows the top 20 strongest collocates of the *BE like* construction, ranked by mutual information (MI) score, where an MI of 3 or more is considered significant. The underlined words in the table signify the identified collocate in the COCA, and only phrases with a frequency of 10 or more are included in the table.

Here, we can see that lexicalized similes appear to be most commonly used in spoken English, and that they are used very infrequently in academic writing. Adding further weight to the spoken results, an analysis of the lexicalized similes in the news section revealed that many of these occurrences involve reported speech. These results are not surprising, because it would make sense for lexicalized phrases in general to be used most in spoken English as they facilitate the speed of language production and processing.

Table 5.1: Frequency of lexicalized *BE like* similes in the five sections of the COCA

Lexicalized <i>BE like</i> simile	Freq.	MI	Spok.	Fic.	Mag.	News	Acad.
Catnip	12	5.09	4	0	5	1	0
<u>jigsaw</u> puzzle	34	4.87	8	7	5	9	5
the <u>Pied Piper</u>	14	4.73	5	2	2	5	0
<u>Rorschach</u> test	11	4.71	7	0	2	3	1
(Russian) <u>roulette</u>	26	4.63	1	4	2	4	2
<u>Groundhog Day</u>	16	4.47	11	0	4	4	0
sponges	21	4.25	7	2	6	3	3
putty	13	4.13	7	3	3	0	0
the proverbial . . .	28	4.07	4	6	8	6	3
herding (cats)	17	4.01	8	2	4	3	1
kaleidoscope	12	3.95	3	5	1	1	0
<u>caged</u> (animal)	15	3.94	2	6	5	4	0
<i>déjà vu</i>	11	3.80	7	1	0	2	1
sandpaper	11	3.69	1	3	6	3	0
sauna	11	3.67	1	5	3	1	0
roller <u>coaster</u>	33	3.64	16	6	4	6	1*
<u>rearranging</u>	10	3.59	3	1	0	5	1
(chairs on the <i>Titanic</i>)							
cockroaches	10	3.55	3	2	3	2	0
<u>Jeckyll</u> (and Hyde)	13	3.47	3	2	4	4	0
Houdini	10	3.46	2	4	4	0	0
			103	61	71	66	17

Wikberg (2008, 130) noted that many lexicalized similes seldom require an elaboration or explanation of their grounds, and this is clearly evident in Examples 5.21 and 5.22. There are, however, several examples appearing with the data as shown in Table 6.1, examples in which, following the use of a lexicalized simile, the grounds are explained or elaborated on. The lexicalized simile *BE like sponges* is often followed by an elaboration that includes *absorb* or *soak up*, and likely most noteworthy is *BE like Groundhog Day*. This lexicalized simile is referring to a well-known movie's premise, so that it is used when a situation or event seems to keep starting over, starting in the same way each time. Several instances of this simile, as found in the data, explicitly state the grounds (the basic premise of the movie, as found in Example 5.24) and even a full synopsis of the movie (see Example 5.25), apparently for the benefit of those who are not familiar with it.

- (5.21) In other words, Kingpin is going to **be like catnip** to the serious gamer. (NEWS[Atlanta])
- (5.22) She flipped on the AC and adjusted the vents. The interior **was like a sauna**. She checked the dashboard. (FIC[Bk:TwoRoadsHome])
- (5.23) But **they're like sponges**. They absorb everything we tell them.
- (5.24) It **was like Groundhog Day** all over again. He said the same things and the same results we're seeing today. (SPOK[CNNLIVEEVENT/S])
- (5.25) Finally somebody said, this **is like the movie Groundhog Day**. Bill Murray plays a character who relives one day of his life over and over. He has to figure out how to get to the next day of his life, so he tries all sorts of things. He begins by learning simple things, like remembering not to step in a puddle on the street. Eventually he learns to stop living just in his mind, to listen to his heart and soul. (MAG[Fortune])

Finally, another lexicalized simile worthy of further discussion is the use of the expression *BE like the proverbial* followed by a well-known proverb or idiom. This self-referential expression is used when the speaker or writer wants to emphasize that the comparison they are making is based on a well-known saying. It is a relatively frequent lexicalized frame, and one that is found in all five sections of the COCA, as shown in Examples 5.26–5.30 below. The most common proverb found to occur with this construction is the proverb about the blind men examining an elephant: this occurs in five out of the 28 instances taken from the data. It should also be noted that three of these five examples include elaborations or explanations of the grounds.

- (5.26) He's **like** the proverbial tornado. You know he's going to do a lot of damage but you don't quite know which path he's going to take.
(SPOK[PBS_NewsHour])
- (5.27) The first column **was like** riding the proverbial bicycle; you may be shaky, but you never forget.
(FIC[Bk:LoudClear])
- (5.28) "When they added even more irrigation canals in the 1960's it **was like** the proverbial straw that broke the camel's back," he said. (MAG[NatGeog])
- (5.29) As for Haiti, it seems, an invasion will **be like** throwing the proverbial rock into the pond. Even if the splash doesn't get you, the ripples could go on forever. (NEWS[WashPost])
- (5.30) In many respects, we **are like** the proverbial blind men trying to reconstruct the elephant.
(ACAD[EnvirAffairs])

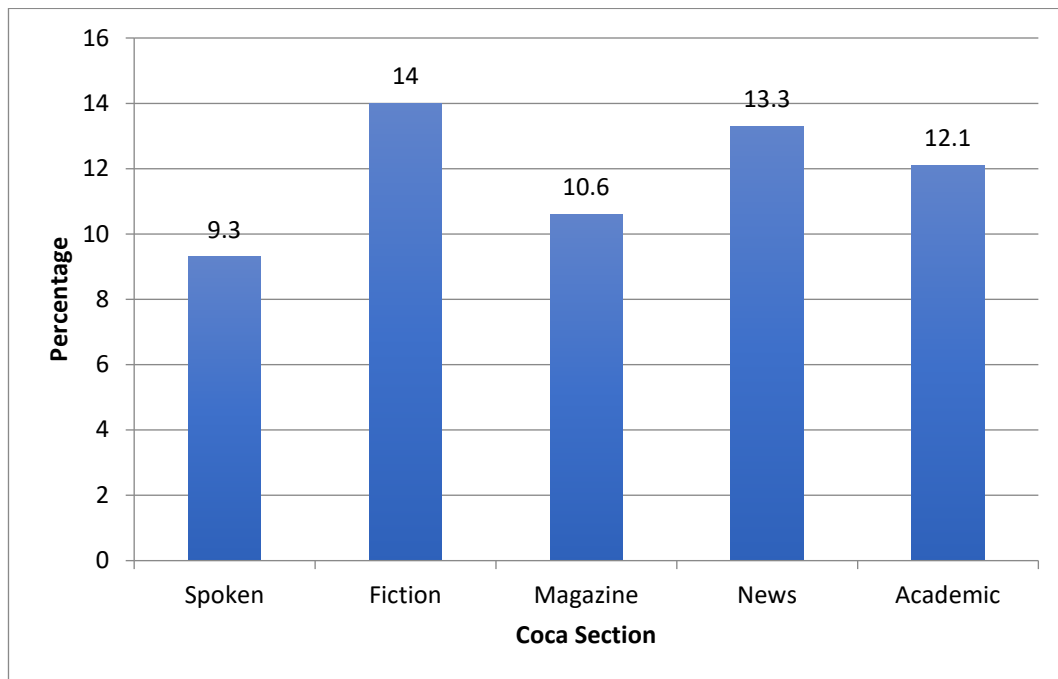
5.3.4: Metaphorical Similes

As mentioned in Section 3.5.4, a metaphorical simile is a simile with a metaphor embedded in the vehicle of the comparison. The data from the corpus

sample, presented in Figure 5.4, shows that this type of simile is relatively uncommon across all five genres of the COCA.

Examples of metaphorical similes from each of the five COCA sections are then provided in Examples 5.31 to 5.35. A small number of the metaphorical similes found in the data sample were identified as conceptual metaphors, such as we see in Example 5.31, which represents the SEEKING = HUNTING conceptual metaphor.

Figure 5.4: Percentage of metaphorical similes by section of the COCA



- (5.31) Greensfelder says that for die-hard collectors like himself, these past two weeks have **been like** open hunting season. (SPOK[NPR-ATC])
- (5.32) It **was like** her face was up in the moon lookin' down at me. (FIC[Triquarterly])
- (5.33) It's **like** dessert for my skin. (MAG[Cosmopolitan])
- (5.34) "Playing Lena **was like** walking a tightrope," said Ms. Bassett. (NEWS[NYTimes])
- (5.35) Talking with Ned **is like** falling down the rabbit hole into a Leibnizian universe. (ACAD[AmerScholar])

5.4: Target–Vehicle Dissimilarity Across Genres

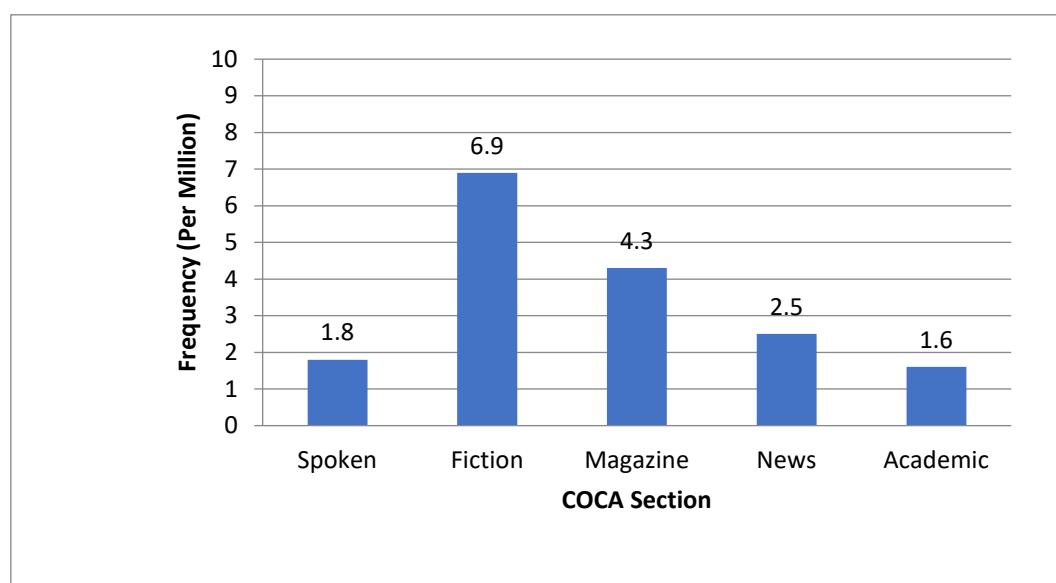
In order to better facilitate examining genre and patterns of dissimilarity, similar to the work of Moon (2009), I used the narrowed-down target pattern of *NOUN BE like NOUN*, using a 100-concordance-line sample from each of the five COCA sections. After removing all examples of non-simile usage, the data were sorted into patterns of target–vehicle dissimilarity for each section.

5.4.1: Frequency Data for *NOUN BE Like NOUN*

5.4.1.1: Frequency of Simile Usage with *NOUN BE Like NOUN*

The *NOUN BE like NOUN* data sets produced samples with much higher proportions of true simile usage than the initial *BE like* data sets ranging from 80 percent in spoken English to 99 percent in academic English. The overall frequencies of *NOUN BE like NOUN* simile usage per COCA section are shown in Figure 5.5.

Figure 5.5: Percentage of true simile usage in 100-concordance-line samples of *NOUN BE like NOUN* from each of the five sections in the COCA



This reasserts the previous finding that academic English uses more nominal-predicative vehicles than the other genres; additionally, it shows that the targets of these similes also tend to be nominal. Another reason for these results is likely due to the large proportion (in the spoken, fiction, magazine, and news sections) of pronouns used in the place of the target position. By contrast, academic English tends to use fewer pronouns, and this is likely due to its more formal nature.

Comparative analyses of the target–vehicle dissimilarity patterns used in the *NOUN BE like NOUN* construction revealed both common and differentiating features with regard to shared and characteristic patterns of dissimilarity among the five sections of the COCA. I will outline these in the next few sub-subsections.

5.4.1.2: Most Frequent Target–Vehicle Dissimilarity Patterns

Thus far, the discussion has shown that, in relation to both frequency of usage and simile type, there are notable differences in *BE like* simile usage across different genres. And this analysis, showing patterns of dissimilarity, has revealed that similes in all five sections of the COCA often share a large proportion of subject matter for both targets and vehicles. During the initial sorting and ad hoc categorization of the *NOUN BE like NOUN* data samples, I identified a total of 27 unique categories of targets and vehicles. (The full list of these categories can be seen in Appendix 2.) However, the majority of simile targets, in all five sections, fell into only five categories: *people*, *body part*, *action*, *abstract*, and *object*. In addition, the majority of the vehicles for these similes came from only ten of the 27 total categories: *people*, *object*, *nature*, *food*, *animal*, *action*, *abstract*, *event*, *sport*, and *family*. This suggests that, regardless of the genre of their discourse, people tend to use *BE like* similes to refer to the same general topics. Table 5.2 shows the

frequency data for the most common target–vehicle dissimilarity patterns in each of the five COCA sections.¹ These findings correspond to those of Hanks (2005).

Table 5.2: Frequency of the most common target–vehicle dissimilarity patterns from 100-line samples of *NOUN BE like NOUN* similes from each section of the COCA

Section	Target–Vehicle Dissimilarity Pattern							
Spoken	people =	11	abstract =	4				
	people		abstract					
	people =		abstract =					
Fiction	object	3	event	3				
			abstract =	3				
	body part =	14	people =	5				
	object		people					
	body part =	13	people =	2				
	nature		animal					
	body part =	4						
Magazine	food							
	body part =	3						
	animal							
	body part =	2						
News	event							
	people =	10	action =	8	abstract =	3	object =	8
	people		action		abstract		object	
	people =	2	action =	3	abstract =			
Academic	object		sport	2	sport			
			action =					
	family		object					
Academic	people =	11	action =	4	object =	2		
	people		action		object			
	people =	3						
	nature							
Academic	people =	14	abstract =	3				
	people		abstract					
	people =	7	abstract =	2				
	animal		nature					
Academic	people =	4						
	nature							

1. Only simile targets recurring more than ten times are included in these data, and only simile vehicles that occur more than once are reported.

5.4.2: Patterns of Target–Vehicle Dissimilarity

In his examination of the collocational behaviour of semantically related words, Walker (2011, 297–298) made a distinction between what he termed shared and characteristic collocates. Shared collocates are collocates that are associated with all of the items in a group, and characteristic collocates are defined as those that are associated with only one of the items in the group. Walker also noted that shared collocates are more frequent than characteristic collocates, and this understanding can be very useful for comparative analysis. Here, I will apply this same shared/characteristic analysis to the target–vehicle dissimilarity patterns of *BE like* similes found in the COCA, attempting to gain insight on differences in simile usage across genre.

Table 5.3 presents the recurring target–vehicle dissimilarity patterns from the *NOUN BE like NOUN* data, and it shows whether they appear in more than one genre (shared) or if they mainly occur in only one genre (characteristic).

Table 5.3: Shared and characteristic target–vehicle dissimilarity patterns for the *NOUN BE like NOUN* construction in the five sections of the COCA. Red x emphasizes that the dissimilarity pattern recurs in only one section and is therefore characteristic of that genre only. *S*, *F*, *M*, *N*, and *A* refer to the spoken, fiction, magazine, news, and academic sections of the COCA.

Dissimilarity Pattern	S	F	M	N	A
people = people	x	x	x	x	x
people = object	x		x		
people = animal		x			x
people = family				x	
people = nature					x
body part = object		x			
body part = nature		x			
body part = food		x			
body part = animal		x			
body part = event		x			
abstract = abstract	x		x		x
abstract = event	x				
abstract = object	x				
abstract = sport			x		
abstract = nature			x		x
action = action			x	x	
action = sport			x		
object = object			x	x	

5.4.2.1: Shared Target–Vehicle Dissimilarity Patterns

Similes About People. The data represented in the samples reveal that similes describing people are of the most common type found in the COCA overall. They account for nearly one-third of all *NOUN BE like NOUN* similes in both the academic and spoken sections, and they make up 23 percent, 16 percent, and 12 percent in the newspaper, magazine, and fiction sections respectively. In all five sections, however, the *people = people* dissimilarity pattern is the dominant pattern, accounting for approximately half of all people-related similes in academic, fiction, and spoken English. This pattern is even more dominant in the newspaper and magazine sections, with nearly three-quarters of similes with people for targets having a different group of people as the characteristic vehicle. Examples of this common *people =*

people pattern of dissimilarity, for each of the five sections of the COCA are provided in Examples 5.36–5.40 below:

- (5.36) Thornburgh and his crew **were like** artists
shaping the landscape by hand. (NEWS[Pittsburg])
- (5.37) Parents often say that pediatricians **are like**
members of the family who watch the children
grow up and change. (MAG[Parenting])
- (5.38) Actors who submit to the knife **are like** athletes
who inject steroids, fakes who should be
disqualified from the race . . .
(ACAD[Southwest review])
- (5.39) The colonists **were like** kids who'd been let out
of school—they wandered fields and dens for the
sheer pleasure of stretching their legs—in a
way, for the sheer pleasure of setting foot on a
strange planet, the whole reason they'd come to
Ulla in the first place. (FIC[Analog])
- (5.40) Venture capitalists **are like** hitchhikers,
hitchhikers with credit cards. And as long as
you take them to where they want to go, they
will help you pay for the gas.
(SPOK[NPR:HowIBuiltThis])

In addition to the frequent *people* = *people* pattern of dissimilarity, several secondary vehicle trends were found—in all sections except for the newspaper section. In both the spoken section and the magazine section, the *people* = *object* pattern is the second most frequent people-related simile pattern, making up 17 percent and 13 percent of the sample in those sections, respectively. This pattern is shown in Examples 5.41 and 5.42.

- (5.41) So, I mean, but you—most performers **are like**
sponges and we all pick up from one another and
you learn how to do certain things vocally and
also interpretively, how to interpret something
in a certain way. (SPOK[NPR_FreshAir])
- (5.42) In a time when many educators are wringing their
hands over the future of astronomy in schools,
these three young women **are like** candles in the
darkness. (MAG[SkyTelescope])

People were compared to animals in 27 percent of the fiction similes, and in 18 percent of the academic similes.

- (5.43) Further, he explained that the rural people **are like** pigs, who work to enrich the urban people, or the dogs. (ACAD[HumanOrganization])
- (5.44) The squatters **are like** roaches in a dark kitchen at night, scrambling as the lights turn on. (FIC[Mov:Bamboozled])

Finally, in addition to *people = people* and *people=animal*, 13 percent of the people-related similes in the academic section use elements of nature as a vehicle, as seen in Example 5.45.

- (5.45) The young girl in the house of her mother **is like** seed in fertile ground. (ACAD[Symposium])

This high proportion of similes about people is not entirely surprising. Humans are highly social animals, and we spend a lot of time talking about people and describing people. The results reported here are likely a reflection of this human tendency.

Similes about abstract things. Abstract nouns supplied the second most common simile target in the data, accounting for over 10 percent of the samples in the spoken, magazine, and academic sections. In the spoken section, the proportion was 15 percent; in both the magazine and academic sections, it was 11 percent. While each of these sections has characteristic vehicle patterns for a portion of the abstract similes (discussed separately in 4.2), the *abstract = abstract* pattern of dissimilarity was the most frequent. Moreover, it was common to all three sections, in nearly identical proportions. Four out of 15 (26 percent) of the abstract-related similes in the spoken section used an abstract noun for the vehicle, and the magazine and academic sections each had 27 percent, or three out of 11. Examples 5.46–5.48 show how these patterns are used in these sections.

- (5.46) Remember, that's what Hillary Clinton–marriage

- is like** slavery, and an Indian reserve—have they taught you that yet? (SPOK[Ind_Limbaugh[]])
- (5.47) Our relationship **is like** politics. You give me this; I'll give you that (MAG[HarpersBazaar])
- (5.48) The scientist's detachment through thought from the dreariness and troubles of merely human life **is like** pantheism in being self-denial through self-divination. (ACAD[PerspPolSci])

It is interesting that abstract nouns are used to describe other abstract nouns when it would seem more helpful to use a concrete noun if the goal is to help someone understand a concept. In the examples given above, however, it can be argued that the speaker or writer is giving an opinion rather than an explanation of a concept. This notion, of simile usage for stating opinion, will be further discussed in Section 5.5.2.

In addition to *abstract = abstract*, the pattern of dissimilarity in *abstract = nature* was found to recur in both the spoken section and the academic section. Accounting for 13 percent and 18 percent of the abstract-based similes in those sections respectively, it is represented in Examples 5.49 and 5.50.

- (5.49) And the caption reads, Dreams **are like** rainbows. Only idiots chase them. (SPOK[NPR_Dayreak])
- (5.50) In the Himalayas, mind **is like** sky, wandering thoughts like clouds, and the body **like** Mount Everest, the White Elephant goddess, nurturing earth. (ACAD[IndepSchool])

Example 5.50 is noteworthy, not only for demonstrating the *abstract = nature* dissimilarity pattern, but also because it contains two *like* similes in sequence, with the second simile (*body part = nature*) containing an embedded metaphor.

Similes about objects. Similes describing physical objects with the *NOUN BE like NOUN* construction accounted for over 10 percent of the samples in both the magazine section and the newspaper section of the

COCA, at 12 percent and 11 percent respectively. The vehicles of these similes are almost always other objects. Of the 12 *NOUN BE like NOUN* similes with an object as the target in the magazine section, eight of them, or 67 percent, fall into the *object = object* pattern of dissimilarity. This figure is even higher in the news section, with nine of the 11 (82 percent) of the object-related similes showing *object = object* patterns.

Example 5.51 (found in the magazine section) and 5.52 (from the newspaper section) are representative of the *object = object* patterns of dissimilarity. Here we can see the same pattern being used for very different reasons: the first describes a product for the purpose of marketing, and the second describes an event.

- (5.51) Foldable kayaks **are like** jigsaw puzzles: They take a long time to assemble and provide lasting frustration. The Oru is the first folding performance kayak that can be assembled in less than five minutes. (MAG[PopScience])
- (5.52) Those who got hit went under. Their packs **were like** anchors, and they drowned. (NEWS[Houston])

Similes about actions. *NOUN BE like NOUN* similes used to describe actions also occurred in the magazine and newspaper sections, this time accounting for 13 percent and 10 percent of the samples respectively. Actions make up the majority of vehicles in these similes, for both sections, with the *action = action* pattern of dissimilarity accounting for 62 percent and 40 percent of the action similes respectively. The remaining 60 percent of action-based similes in the newspaper section consisted of various dissimilarity patterns occurring only once and were therefore omitted from this discussion. Two more action-based patterns of dissimilarity were found in the magazine section, *action = sport* and *action = object*, but these occurred

only in this section and will therefore be covered below, in my discussion of characteristic dissimilarity patterns (4.2.3).

The majority of action-related similes follow with actions for vehicles, often analogously, with this sort of dissimilarity pattern occurring in 62 percent of the sample (5.53). The next most frequent vehicle group is related to sport, accounting for 23 percent of the data (5.54).

- (5.53) Remember character building **is like** house building: It's supposed to be tough, otherwise everyone would do it. (MAG[MotherEarth])
- (5.54) Making a movie **is like** baseball in that way. It's a lot of waiting around, thinking about what you're getting ready to go do. And then when you do it, it's a spring, full out. (MAG[Esquire])

5.4.2.2: Characteristic Target–Vehicle Dissimilarity Patterns

Spoken. As previously noted, the highest proportion of *NOUN BE like NOUN* similes with abstract noun targets appeared in the spoken section of the COCA, with the *abstract = abstract* pattern of dissimilarity being the most frequent. This was shared with the magazine and academic sections, and *abstract = nature* was also shared with the academic section. Two additional abstract noun-based dissimilarity patterns, however, were found to be characteristic of the spoken section: *abstract = object* and *abstract = time*.

The *abstract = object* pattern of dissimilarity occurred three times in the sample, or in 20 percent of all the abstract-based similes. Two of these use the lexicalized, and metaphorical, vehicle of *gold* as seen in Example 5.55.

- (5.55) And negative coverage of a certain type **is like** gold for these candidates. (SPOK[ABC:ThisWeek])

The *abstract = time* pattern of dissimilarity occurred twice, that is, in 13 percent of the sample, and in both cases the lexicalized vehicle *day and night* is used.

- (5.56) The difference between this Eid and previous Eids
under the Taliban **is like** day and night.
(SPOK[NPR_Sunday])

Fiction. The most prominent genre-specific pattern of dissimilarity in the spoken section of the COCA was found to involve simile targets focused on body parts. The most common vehicles for body part similes in fiction involve physical objects, accounting for 38 percent of all occurrences (as shown in Example 5.57), and elements of nature accounting for 35 percent, as shown in Example 5.58. In addition, 11 percent of the body part–related similes in this section use food as the vehicle (as in Example 5.59). Also characteristic to fiction, although less frequent, are the body part = animal, and body part = event dissimilarity patterns (Examples 5.60 and 5.61 respectively).

- (5.57) I work till my eyes **are like** cotton.
(FIC[Antipodes])
- (5.58) Her bare, pale legs **were like** twigs stripped of
their bark and he remembered how he used to do
that, long ago, too. (FIC[NewYorker])
- (5.59) Her tongue was stained orange, and her hair **was
like** syrup running down her back.
(FIC[Bk:TwoRivers])
- (5.60) Her face was skull-shadowed, and the veins on
her thin hands **were like** earthworms under wax
paper. (FIC[ArkansasRev])
- (5.61) His smile **was like** Christmas.
(FIC[Bk:NoGood-byes])

Magazine. Two characteristic target–vehicle dissimilarity patterns were identified in the magazine-section data: *action = sport* and *action =*

object. The more frequent *action = action* was also prominent in the newspaper section, as discussed in 5.4.1.2.

Of the 13 *NOUN BE like NOUN* similes with action-based targets, three (or 23 percent) used a sport for the vehicle (Example 5.62), and two (15 percent) used an object for the vehicle (Example 5.63).

(5.62) Making a movie **is like** baseball in that way.
It's a lot of waiting around, thinking about
what you're getting ready to go do.
(MAG[Esquire])

(5.63) "Really resting for the first week **is like** money
in the bank," says midwife Kathy Wilson.
(MAG[TodaysParent])

News. All three of the main simile target groups in the newspaper section—people, action, and object—are shared with other sections of the COCA. However, I identified one subset of the *people = people* pattern of dissimilarity in the data as particularly characteristic of the newspaper section. Out of the 22 similes with people-related targets, there were five occurrences of *people = family*, accounting for 23 percent of the *people = people* similes, as shown in Example 5.64.

(5.64) "My friends **are like** family," Davenport says.
(NEWS[Atlanta])

Academic. Only one characteristic target–vehicle dissimilarity pattern was identified in the academic section, which is a subset of target groups shared with other sections. The *people = nature* pattern occurred four times out of the 30 people-based similes, accounting for 13 percent of that group (Example 5.65).

(5.65) "Your ancestors **are like** streams," her mother
once told her, "their waters may spill from many
different lands, but finally, in you, they flow
together. (ACAD[Writer])

5.5: Rationale Behind Simile Usage

5.5.1: Describing the Unfamiliar

In Moon's (2011) textual analysis of simile usage, the main function of similes in travel writing and fiction were to present the unfamiliar in familiar terms, and to present the familiar in unfamiliar terms. In her conclusion, she wrote as follows:

This paper set out to look at how similes work through linking together dissimilar things, and what the nature of the dissimilarities, the patterning of the “rival realities” within similes, reveals about the texts in which they occur. In the explorers' texts discussed here, similes mainly work by recasting the unfamiliar in terms of the familiar, while those in Conrad's *Heart of Darkness* more or less do the reverse. Perhaps this reflects some wider generic distinction between fiction and nonfiction, though it would be much too large a claim to make here. Even where similes appear simply to explain and elucidate, their dissimilarities are meaningful and may have ideological significance. (153)

The data from this study do not support Moon's supposition that fiction and nonfiction differ in this way concerning simile usage. Known to unknown patterns of dissimilarity are not commonly found in the COCA, and this suggests that such patterns mark a literary device unique to Conrad's particular work (especially *Heart of Darkness*), without reflecting fiction as a whole. Rather, the results in this chapter suggest that the main purpose of using *BE like* similes is the same across all genres—that is, to explain the unfamiliar in familiar terms respectively.

However, this is a generalization; and closer examination of the data to some degree suggests that there are several different and more specific reasons for using *BE like* similes. These other reasons may have their nuances, but they generally fall under the same pattern of explaining the unfamiliar in familiar terms. Furthermore, with regard to the similes used by

Conrad, it was pointed out to me by my editor (credited in my acknowledgements) that Conrad's unique use of simile may have been a deliberate literary device developed specifically in *heart of Darkness*. He notes that

the novella stands out, precisely for its mode of exploration, but a story of exploration in literary terms, intended to challenge the reader to try and see what is in the heart of darkness, or where that heart is: in Africa? Europe? Humanity? Kurtz alone? (MacDonald, personal communication, 2020)

It seems possible, then, that Conrad's use of familiar = unfamiliar dissimilarity patterns was deliberate and may even be unique to this one work and others that Conrad adapted from his travel journals. It is also worth noting here that English was not Conrad's native language. While recognized as being one of the very best modernist writers of English-language fiction, he did not grow up speaking English (MacDonald, personal communication, 2020). Of course, this speculation runs in the direction mentioned by Moon in her paper—far beyond the scope of the present study.

5.5.2: Creativity and Opinion

A large proportion of the similes in all five of the COCA's subcopora, on the other hand, would seem to describe the familiar in familiar terms. These are always nominal similes, and they are followed by a specific explanation of the grounds. They often seem to function for the purpose of creative description, or for that of offering opinions, as in Examples 5.66 and 5.67.

- (5.66) Bankers **are like** trapeze artists. The greater the net you put under them, the greater the risk they'll be willing to take. (SPOK[ABC_20/20])
- (5.67) Electricity **is like** cash in the desert. Everybody wants it. Everybody needs it. And I've found myself willing to trade in all I have to get it. (NEWS[AssocPress])

Table 5.4 shows the percentages of similes where both the target and the vehicle belong to the same general category. In many of these instances, it seems that the general pattern is describing something that is known in terms of something else that is known, and that it could be construed as giving an opinion or view of the target. However, offering an opinion can be seen as explaining the unknown. In Example 5.66, most people know what a banker is and what a trapeze artist is, yet the simile is being used to give a (seemingly low) opinion, unknown to the audience until the speaker or writer reveals it.

Table 5.4: Percentage of *BE like* similes with target and vehicle belonging to the same category

Section	Percentage
Spoken	32
Fiction	22
Magazine	44
Newspaper	40
Academic	32

5.5.3: Elaboration

Another reason for using *BE like* similes is for elaboration, or for clarification of an already-given description. This is evident in the data, as the most frequent item to precede *BE like* is the pronoun *it*. This accounts for nearly half of all the similes in the fiction, magazine, and news sections, and for 60 percent of the *BE like* similes in spoken English. Academic English has a significantly lower proportion of *it BE like* patterns, with only 24 percent of its occurrences having this pattern. In all these instances of *it BE like*, the target has already been mentioned, and the simile is an elaboration or clarification. Examples of *BE like* similes used for elaboration are shown in 5.68 to 5.70 below.

- (5.68) He's perfect—too perfect. **It's like** Apple was parodying its own image while also cementing it. (SPOK[NPR_Daybreak])
- (5.69) It made him feel alive. He said **it was like** being a soldier again. (FIC[VirginiaQRev])
- (5.70) Exposure to excellent writing strengthens children's grasp of vocabulary, sentence structure, narrative development. **It's like** the language seeps into them. (MAG[TodaysParent])

5.6: The Complexities of Simile

In Chapter 3, I pointed out that while simile has been examined in some detail in the literature, it is rarely the main focus of study. Rather, descriptions of simile are most commonly included in discussions of metaphor, in which the two phenomena are compared, a simile is often dismissed as a subset, or lower form, of metaphor (Hanks, 1; Veale and Hao 2007, 683). This incongruence seems misguided, since there are claims that simile occurs just as often as metaphor (Fadaee 2011, 22). Furthermore, with the notable exceptions of studies by Hanks, Moon, and Wikberg, there are relatively few corpus-based phraseological studies of simile in the literature. Even fewer corpus-based studies on simile examine genre differences. Furthermore, the examples provided in many studies are simplified and removed from their context (Wikberg, 129).

While there are competing models of metaphor, both the categorization and career of metaphor models recognize simile as its own separate and independent phenomenon from metaphor. These models do not treat it a lower, or simpler, form of metaphor. My studies for this thesis have already begun to show that simile is a complex and worthy subject of investigation in and of itself, without the need for a comparative analysis with metaphor.

In this chapter, I have shown how the frequency of simile usage varies across sections of the COCA. This pattern suggests that there is a corresponding variation among their respective registers. In addition, I have shown that similes of different types are favoured in different genres. And, while the majority of similes can be categorized into a small number of very common patterns (Hanks), there are clear differences among genres in how they favour or disfavour specific patterns. Finally, through corpus-based speculation, I have attempted to explain the rationale and motivation behind *BE like* simile usage across the different genres, and I have suggested that similes are often used for giving opinions and for elaboration.

Meanwhile, the main functions of simile usage discussed in the literature centre around explanation and description, including the expression of emotion. Other functions are said to include making descriptions more vivid and entertaining. When used for general explanation, similes most commonly describe the unknown, or the unfamiliar, in familiar terms (Moon, 142). When used to describe personal feelings, Hanks (9) claimed that *like* similes are “more or less conventional” as they “play an important role in presenting the inner and unknowable feelings of an individual to the outside world.” He provided several examples of the *FEEL like* simile pattern, and he noted that this construction is often used by fiction writers to describe the feelings of their characters. Previous studies have identified several notable differences in how similes are used in different genres. While similes are likely found in most genres, if not all (including prose, poetry, and conversation; Fadaee 2011, 22), previous studies have shown that similes are most commonly found in fiction (Wikberg, 129). This chapter has confirmed that simile is frequent across all genres, and it has provided further insight into the specifics of simile usage. Those specifics include its use for

elaboration and for giving opinions. Finally, the findings presented in this chapter have established that lexicalized similes are more common in spoken English.

6

Results

Main Functions of *VERB Like*6.1: *Like* with Verbs; Perception

In this chapter, I will report on my examination of the second most frequent pattern found to occur with *like*, the *VERB like* construction, in order to identify its most frequent functions, including its main patterns of usage, phrases, and collocates. As I have previously mentioned, *like* is a highly frequent word form, occurring 831,771 times (1,440 words per million) in the 560-million-word Corpus of Contemporary American English (COCA). As a preposition, it is the 73rd most frequent word in the corpus, and one-third of its occurrences consist of the *VERB like* construction. Table 6.1 shows the frequency data and MI scores for the top ten most frequent verbs used with *like* in the COCA.

Table 6.1: Most frequent verbs occurring with *like* in the COCA, with mutual information (MI) values. Verbs in upper case are lemmatized. An MI score of 3.0 or above shows that the collocation has statistical significance, and higher MI scores signify stronger collocations.

Verb	Frequency	MI
LOOK	72,604	5.48
FEEL	48,659	5.88
SOUND	23,175	6.25
SEEM	20,771	5.09
ACT	6,332	4.44
SMELL	3,674	5.52
TASTE	1,913	4.49
WORK	1,815	0.31
shaped	1,764	6.33
TALK	1,743	1.13

While Sinclair (1991) cautioned against lemmatization when conducting phraseological research, my initial inspection of the data showed that the various forms of the verbs in the *VERB like* construction typically do not differ significantly in their usage. The verbs in this study will therefore be lemmatized so that (e.g.) *look like*, *looks like*, *looked like*, and *looking like* all have the same meaning. There is one exception to this lemmatization, however, as can be seen in Table 6.1, that demonstrates Sinclair's point about different forms taking different meanings. The verb *SHAPE* has not been lemmatized because I found it to have different meanings associated with its different forms. Only the past-tense form was found to collocate with *like* in the *VERB like* construction. My examination of concordance lines for the non-collocating forms revealed that, in many instances, the “verb” in the pattern was actually being used as a noun. As shown in Example 6.1, *like* was sometimes used in its *such as* or *for example* sense (*like*_{1.2}), while Example 6.2 shows that it sometimes appeared as part of some other, unrelated phrase.

- (6.1) You get extra points for sculpting in a sassy
 shape **like** a heart or a star. (MAG[Cosmopolitan])
 (6.2) “I’m not in shape **like** you guys,” he claims.
 (MAG[Backpacker])

As previously mentioned in Chapter 3, two types of verbs are found in this grammatical pattern (Wikberg, 2008 138-139): Predicative, *VERB like*, verbs consist of the five main sense verbs (SV) (*LOOK*, *FEEL*, *SOUND*, *SMELL*, & *TASTE*) and the verb *SEEM*, usually followed by a predicate. These six verbs can be grouped together and will be collectively referred to as perception verbs (PV). The adverbial, *VERB like* group, includes any non-PV that is typically followed with an adverbial phrase. These will be referred to as non-PVs.

The data in Table 6.1 show that the vast majority of *VERB like* usage involves four PVs, *LOOK*, *FEEL*, *SOUND*, and *SEEM*. While Wikberg (2008, 138) reported that these PVs account for 52 percent of *VERB like* occurrences in the BNC, their usage was found to be notably higher in the COCA, at 71 percent. The remaining two PVs, *SMELL* and *TASTE*, were significantly lower in frequency compared to the top four PVs.

The remaining non-PVs that occur in the *VERB like* construction, shown in Table 6.1, constitute what Wikberg (2008, 139) referred to as the adverbial subcategory of *VERB like* verbs. (I will refer to them as non-PVs.) She identified over 1,500 verbs that occurred in this adverbial *VERB like* pattern, and she noted that the “vast majority” of these occurred only once or twice in the BNC, with the verbs *BEHAVE* and *ACT* being most common.¹

This finding differs somewhat from the COCA data in Table 6.1, where three of the top ten verbs in the *VERB like* construction are non-PVs that occur more than one thousand times each. Of the three resulting collocations, however, only *ACT like* produces an MI score above three, thus meeting the standard for collocational significance. Based on these findings, it was my decision to focus this part of the study solely on the six PVs *LOOK*, *FEEL*, *SOUND*, *SEEM*, *TASTE*, and *SMELL*. Accordingly, no further analyses will be carried out on the non-PVs in Table 6.1.

In my study for this chapter, I found that when the most frequent sense verbs that pair with *like*—*LOOK*, *FEEL*, and *SOUND*—appear to be

¹. There is a notable difference between corpora here, as data from the COCA (Table 6.1) identify five verbs that occur with *like* more frequently than *BEHAVE* (*WORK*, *shaped*, *TALK*, *DRESS*, and *LIVE*). This discrepancy in frequency information between the COCA and the BNC is likely a reflection of the differences in vocabulary usage between American and British English. Concerning the synonyms *ACT* and *BEHAVE*, British English seems to prefer *BEHAVE like* over *ACT like*, with 4.22 and 2.22 occurrences per million (OPM) respectively in the BNC, whereas the opposite is true in the COCA, where *ACT like* shows 10.96, and *BEHAVE like* shows 3.49 OPM.

polysemous, and for large proportions of their usage, they do not behave as sense verbs at all. It was found that the three most frequent ‘sense’ verbs in the *PV like* construction are often used in the same way as *SEEM like*, that is, for the function of situational evaluation, or what Hanks (2005, 5) referred to as the tentative evaluation of propositions. And when grouped together, the corpus data shows that these four verbal constructions share a considerable range of collocates and phrases. Furthermore, while a cursory examination the remaining two *PV like* constructions suggests that *SMELL like* and *TASTE* are used exclusively for sensory description, or as true sense verbs, a more in-depth analysis of the corpus data reveals that these too can be used for situational evaluation, although much less frequently.

In the following sections of this chapter, I will report on my analyses of the five main sense verbs (SV), *LOOK*, *FEEL*, *SOUND*, *SMELL*, and *TASTE*. Along with these, *SEEM* will round out the wider group of PVs as it was found to be closely related to the three most frequent SVs, thus meriting attention in the scope of this study to determine the most frequent functions of the *VERB like* construction. Sections 6.3–6.8 will discuss each of the six *PV like* constructions in detail, after first examining the phraseology shown to be common to the six main verbs in the *PV like* construction in Section 6.2 below.

6.2: *PV Like* Phraseology

6.2.1: *PV Like* Patterning

The most common grammatical pattern associated with all six *PV like* phrases is *NOUN VERB like NOUN*. Hunston, Francis, and Manning (1996), in their comprehensive verb pattern grammar (notated *V like n*), state that verbs in this pattern are all link verbs and “used to indicate how someone

or something seems” (1996: 209). As can be seen in Examples 6.3–6.8 below, this *PV like* pattern is mostly used for descriptions of sensory experience, taking the form of both figurative comparisons (true similes) as shown in Examples 6.3, 6.4, and 6.7, and literal comparisons, as shown in Example 6.8. However, this same pattern can also be used for situational evaluation, as shown in Examples 6.5 and 6.6.

- (6.3) You **look like** a whale sitting on the beach.
(SPOK[ABC_PrimeTime])
- (6.4) The hand gave Christina the creeps, and when she shook it, the hot dry skin **felt like** a reptile’s.
(FIC[Bk:Fire])
- (6.5) “He suggested it and it **sounded like** a good idea,” said Jackson-Portwood.
(NEWS[DetroitFreePress])
- (6.6) Everything **seems like** a dream. (SPOK[NPR_ATC])
- (6.7) His breath **smelled like** a blighted river.
(FIC[NewYorker])
- (6.8) If the liquid tastes like vinegar, it’s ready.
(MAG[MotherEarth])

In addition to nouns and noun phrases, *PV like* also takes full clauses (referred to as clausal similes by Wikberg (2008)), although much less frequently. And while there are instances of *NOUN PV like CLAUSE* with all six PV phrases as seen in Examples 6.9–6.14, they are most common with *FEEL like*.

- (6.9) Well, it **looks like** it’s going to be a long one.
(SPOK[ABC_ThisWeek])
- (6.10) When you do, it **feels like** you’re being lit on fire. (FIC[Bk:ThunderRain])
- (6.11) It **sounds like** life is really on the upswing for you. (MAG[Shape])
- (6.12) He **seemed like** it’s normal.
(SPOK[ABC_PrimeTime])
- (6.13) He **smelled like** he spent most of his time in smoking rooms, (FIC[Analog])
- (6.14) She **tastes like** she just ate a tootsie pop,

(FIC[IowaRev])

Hunston, Francis, and Manning (1996: 452) refer to this pattern as “*V as if/V as though*” and note that this usage of *like* is found in informal English. The most common function associated with this pattern in the data is situational evaluation (Examples 6.15–6.18), but it can also be used for the description of sensory experience (Examples 6.19 and 6.20).

(6.15) Well, it **looks like** it’s going to be a long one.
(SPOK[ABC_ThisWeek])

(6.16) “I feel like I definitely could do it,” Johnson
said. (NEWS[NewYorkPost])

(6.17) It **sounds like** you are on the right track.
(MAG[ChildDigest])

(6.18) He **seemed like** it’s normal.
(SPOK[ABC_PrimeTime])

(6.19) He **smelled like** he spent most of his time in
Smoking rooms, (FIC[Analog])

(6.20) She **tastes like** she just ate a tootsie pop,
(FIC[IowaRev])

Several extended patterns were also found to be shared among the six *PV like* constructions, with the most frequent being *make NOUN PV like NOUN/CLAUSE*. This occurs 6,307 times in the COCA, with up to four words between *make* and the PV. This pattern seems to work with both of the core meaning groups: situational evaluation usage can be seen in Examples 6.21, 6.22, and 6.24, while descriptions of sensory experience are apparent in Examples 6.23, 6.25, and 6.26.

(6.21) The skin on his face was darker than his scalp
or jaw, **making it look like** the man had just
shaved off a beard and thick, wild hair.
(FIC[Bk:OrphanMastersSon])

(6.22) It **makes me feel like** a bum. (SPOK[ABC_20/20])

(6.23) It **made his voice sound like** a robot’s.
(NEWS[SanFranChron])

(6.24) He **made it seem like** a business, or like a
partnership. (SPOK[NPR_TellMore])

- (6.25) He used to bake these meringue cookies that **made our apartment smell like** a chocolate factory.
(FIC[Read])
- (6.26) It's like they injected it with something to **make it taste like** fast food. (MAG[Atlantic])

Another extended pattern shared by all six *PV like* constructions is *PV like a cross between NOUN and NOUN*. This is predominantly associated with the description of sensory experience as shown in the examples below.

- (6.27) He **looks like a cross between** Colonel Sanders and Colonel Kurtz in Apocalypse Now.
(MAG[Newsweek])
- (6.28) Sting was cranking a hurdy-gurdy, which **sounds like a cross between** bag-pipes and a fiddle.
(MAG[People])
- (6.29) Monstera, which **tastes like a cross between** the pineapple and the banana, made a nice main course. (NEWS[WashPost])
- (6.30) The Calistoga Inn **feels like a cross between** an Old West saloon and a grandmother's house.
(NEWS[SanFranChron])
- (6.31) a cake artist who "always **smelled like a cross between** a cinnamon stick and a whiff of Old Spice." (NEWS[Denver])
- (6.32) It did **seem like a cross between** a hockey puck and a golf ball. (FIC[InternalMedic])

This construction, which Hanks (2005) referred to as a double comparison, occurs 142 times in the COCA and, like many similes, functions "not to create a precise factual statement, but rather to activate the imagination on the basis of vague possibilities—and to entertain" (9).

6.2.2: *PV Like* Phrases

On the phraseological level, the data revealed a number of semi-fixed phrases shared by several of the PVs. For example, the phrase *PV like fun* was found to occur 285 times in the COCA (see Table 6.2), with *SOUND*,

LOOK, *SEEM*, and *FEEL* always belonging to the core meaning group of situational evaluation. Unsurprisingly, there are no instances of *SMELL like fun* or *TASTE like fun* in the COCA, as *SMELL* and *TASTE* are most often used for sensory description.

Table 6.2: Frequency of perception verbs in the *VERB like* construction, according to the COCA, with mutual information (MI) values

<i>PV [like fun]</i>	Frequency
sounds	119
sounded	25
sound	24
looks	44
looked	20
look	19
seemed	9
seem	7
seems	7
feel	6
feels	3
felt	2

Another semi-fixed phrase, *PV like a million bucks/dollars*, occurs in the COCA 126 times. Its several variations are shown in Table 6.3. Since this phrase is used to convey a sensory experience, there are no instances of this phrase occurring with *SEEM like* in the data. Interestingly, however, there are no occurrences of *TASTE like*, and only three occurrences of *SMELL like* in the entire COCA.

Table 6.3: Occurrences of *PV like a million bucks/dollars* in the COCA

Variation	Frequency
<i>LOOK like a million bucks</i>	49
<i>LOOK like a million dollars</i>	16
<i>FEEL like a million bucks</i>	40
<i>FEEL like a million dollars</i>	12
<i>SOUND like a million bucks</i>	4
<i>SMELL like a million bucks</i>	3
<i>look and feel like a million bucks</i>	1
<i>feel like a million sparkling dollars</i>	1
<i>LOOK like a million</i>	4
<i>FEEL like a million</i>	3

6.2.3: Shared Collocates

As previously described in Section 5.4.2, when examining a group of similar lexical items, it is useful to differentiate between their shared and characteristic collocates (Walker 2011, 297–298), and this form of analysis was used to examine the target-vehicle dissimilarity patterns of *BE like* similes. Shared collocates are defined as those collocates associated with all of the items in a group, and characteristic collocates are defined as those that are associated with only one of the items in the group. As previously mentioned, shared collocates are more frequent than characteristic collocates, and identifying an item's characteristic collocates can provide insight into that item's unique qualities. Here, I will examine the shared and characteristic collocates of the six PVs in order to gain insight into their similarities and differences. Table 6.4 shows the shared collocates among the six PVs. They are presented in descending order by frequency, and they are further sorted into the seven semantic groups of conventional similes outlined by Hanks (2005). These data highlight the similarity in usage among the four most frequent *VERB like* phrases (*LOOK like*, *FEEL like*, *SOUND like*, and *SEEM like*). They also show the notable absence of shared collocates, and thus difference in usage, between these four verbal phrases and those formed with the lower-frequency PVs (*SMELL like* and *TASTE like*).

Table 6.4 shows that only a few collocates occur with all four of the top PVs, *LOOK*, *FEEL*, *SOUND*, and *SEEM*, including child/children man/men, fun, and dream. However, we can see strong connections among two pairs of PVs. First, the data shows that *LOOK like* and *FEEL like* share many of the same collocates, especially with the meanings of describing how a person looks and of describing how one feels, and this is worth noting. This data suggests, that generally, we tend to describe how someone other

than us looks, and how we ourselves feel, with many of the same collocates. This is reflected in the data. When looking at the pronouns most often used with *LOOK like* and *FEEL like*, *you/he/she/they LOOK like* occurs 9,928 times compared to 1,641 instances of *I/we LOOK like*. Conversely, there are 13,851 instances of *I/we FEEL like*, and while *you/he/she/they FEEL like* does occur frequently, 6,467 times, the majority of these are interrogative instances relating to situational evaluation, such as we see in Example 6.33.

(6.33) Do you **feel like** people were looking at you
funny? (SPOK[ABC_20/20]).

Table 6.4: *PV like* and its shared collocates

Human Status	L O O K	F E E L	S O U N D	S E E M	S M E L L	T A S T E	Artifact	L O O K	F E E L	S O U N D	S E E M	S M E L L	T A S T E
(a) (little) kids	x	x	x				shit	x	x				
(a) (little) boy/s	x	x	x				the real thing	x	x				
(a) child/children	x	x	x	x			crap	x	x				
(a) (little) girls	x		x				(a) flower/s	x				x	
(a) guy/s	x			x			smoke	x				x	
(a) man/men	x	x	x	x			Event						
(a) baby/ies	x	x					fun	x	x	x	x		
(a) woman/women	x	x	x				a good idea	x		x			
teenagers	x	x					a great idea			x	x		
Human Role							a good thing			x	x		
(a) winner/s	x	x					rain			x		x	
a princess	x	x					(an) hour/s		x		x		
a team	x	x					an eternity		x		x		
(a) tourist/s	x	x					(a) year/s		x		x		
(a) hero/es	x	x					a scene from / out of	x		x			
(a) criminal/s	x	x					a good deal			x	x		
(a) loser/s	x	x					a good way			x	x		
(a) rock star/s	x	x					a contradiction			x	x		
(a) character/s	x	x					trouble	x		x			
(a) member/s	x	x					a good plan			x	x		
(a) king/s	x	x					Christmas	x	x		x		
(a) victim/s	x	x					(a) death		x			x	
a queen	x	x					a dream	x	x	x	x		
Human Attributes							a lifetime		x		x		
a man who/with/in	x		x				the right thing to do		x		x		
(a) fool/s	x	x					a long time		x		x		

Human Status	L O O K	F E E L	S O U N D	S E E M	S M E L L	T A S T E		Artifact	L O O K	F E E L	S O U N D	S E E M	S M E L L	T A S T E
an idiot	x	x	x					a cartoon	x					
(a) freak/s	x	x						child's play	x			x		
an old man	x	x						Irrealis						
(a) jerk/s	x	x						hell	x	x				
common sense			x	x				(a) ghost/s	x	x				
a different person		x		x				(an) alien/s	x	x				
Animal								magic	x			x		
(a) dog/s	x		x					science fiction			x	x		
(a) fish	x	x			x	x								
(a) bird/s	x	x												
(a) chicken/s	x					x								
(a) rat/s	x	x												
(a) cat/s	x		x											

Next, we can see that *SOUND like* and *SEEM like* have many shared collocates and the majority of these are found in the events category. This shows that *SOUND like* is used in the same way as *SEEM like*, that is, for situational evaluation, more often than *LOOK like*.

Finally, it is also worth noting here, that shared collocates do not always have the same meaning in the *PV like* construction. For example, one shared collocate of *LOOK like* and *FEEL like* from the animal category, *rat*, takes on different meanings depending on the PV. For example, with *LOOK like*, *rat* is often premodified with *drowned* or *wet*, taking on the very specific meaning of being completely soaked, usually from heavy rain, and looking disheveled. When *rat* is paired with *FEEL like*, on the other hand, it is often found in the phrases *caged rat*, *rat in cage*, or *lab rat*, which are used to show feelings of being mistreated. *FEEL like a rat* is also used in the sense that someone feels like a traitor or double-crosser, often associated with the idiom *to smell a rat*.

6.3: *LOOK Like*

LOOK like is by far the most frequent of the *PV like* phrases, occurring 72,604 times in the COCA, and functions as a true sense verb more often than any of the other polysemous PVs. A sample of 500 concordance lines taken from the COCA showed that *LOOK like* is used for the description of visual sense 67 percent of the time. The following three examples illustrate such usage:

- (6.34) This four-seater is really different, **looking like** a 1950s hot rod. (NEWS[Chicago])
- (6.35) That guy right there **looks like** a friend of mine from St. Paul. (SPOK[CBS_SunMorn])
- (6.36) Do you **look like** her as well? (FIC[Bk:DarkRoom])

The other 33 percent of *LOOK like* usage relates to situational evaluation.

- (6.37) It didn't **look like** she was ever going to let Him go. (SPOK[CNN_Chung])
- (6.38) It **looks like** we're really going to do something. (SPOK[PBS_Newshour])
- (6.39) But, right now, it **looks like** we got pretty lucky. (NEWS[Houston])

However, the distinction between visual description and situational evaluation is not clear-cut. This is because there are many instances where the usage could be construed as conveying either visual sense description or situational evaluation, or possibly even both meanings simultaneously. This makes sense of course, as people may tend to base their situation evaluations on what they see—and they would often state these evaluations using *LOOK like*. Examples 6.31–6.34 illustrate this ambiguity.

- (6.40) I **look like** a guy who doesn't know how to ride a bike. (MAG[Bicycling])
- (6.41) It **looked like** a pleasant spot for a picnic. (FIC[Analog])

- (6.42) Zed's bed, while still made, **looks like** it's
been slept in. (FIC[Mov:KillingZoe])
- (6.43) Do I **look like** the protective type?
(FIC[BkSF:Soothsayer])

If we focus only on its use for visual sense description, *LOOK like* can be further categorized into literal comparisons or figurative, true simile, usage. And, here too, the boundary between the two categories is rather fuzzy. There are instances that could be placed in either group, suggesting more of a continuum, rather than a dichotomy, between literal and figurative. Examples 6.44–6.46 illustrate.

- (6.44) Gwen told me I **looked like** a Bolivian housewife.
(FIC[FantasySciFi])
- (6.45) It cost hundreds of dollars for her to **look like**
a doll. (FIC[ContempFic])
- (6.46) It **looked like** a lawn mower ran over the bottom
of the dress. (SPOK[CBS_Early])

6.3.1: *LOOK like* Characteristic Collocates

Being the most frequent PV like phrase, it is not surprising that *LOOK like* is highly phraseological, with the most characteristic collocates of all the PVs. These are shown in Table 6.5, where we can see that *LOOK like* has collocates in all seven of the meaning groups outlined by Hanks (2005). We can also see that it collocates most strongly with words used to describe people (human status, human role, human attributes) and things (artifact).

It is also possible, in most cases, to identify which meaning of *LOOK like* would most likely be associated with each collocate. For example, the most frequent usage of *LOOK like*—the literal description of visual sense—has the most characteristic collocates, with some of the most obvious being *sisters*, *twins*, *brothers*, *models*, *hooker/s*, *wood*, and *marble*.

Table 6.5: Characteristic collocates of *LOOK like* in the COCA with MI scores above three, and frequencies over ten, sorted into the seven main semantic groups outlined by Hanks (2005)

Human Status	Human Role	Human Attribute	Animal	Artifact	Event	Irrealis
(a) person/ people 39 sisters 21 twins 18 brothers 12	(a) movie star/s 50 (a) model/s 51 hooker/s 24 a clown 22 the bad guy/s 20 (a) bum/s 16 Elvis 12 a leader 11 (a) killer/s 11 royalty 10	a guy (who/with/in) 39 (a) genius/es 29 a woman (who/with/in) 22 a dork 17 a wimp 12	(a) duck/s 38 (a) snake/s 22 (a) deer 18 (a) pig/s 18 (an) ant(s) 16 (a) worm/s 13 (a) bear/s 11 (a) horse/s 11 a frog 11	America 91 (a) toy/s 45 blood 44 a war zone 38 a tree 27 a star 24 a city 20 a picture 20 a nail 18 a gun 18 a museum 17 wood 17 a statue 16 junk 14 the moon 13 a car 12 Barbie 12 marble 11 paintings 10	an accident 68 a bomb 26 a bargain 25 a robbery 24 snow 24 a tornado 18 trouble 17 a cartoon 15 a suicide 14 a mess 14	an angel 50 (a) monster/s 22 (a) zombie/s 12

Collocates that would most often be used in similes include *a clown*, *royalty*, *(an) ant/s*, *a war zone*. Animal collocates of *LOOK like* are always figurative when the target of the simile is a person, and usually figurative in most other cases. These include some idiomatic expressions based on animals, such as those found in the following examples:

- (6.47) If it **looks like** a duck, waddles like a duck and quacks like a duck, it is probably a duck.
(NEWS[Denver])
- (6.48) Reese **looks like** a deer caught in the headlights. (FIC[Mov:HotZone])

When *LOOK like* collocates with *man* or *woman* followed by a prepositional phrase, it is usually an indicator of situational evaluation.

- (6.49) He **looked like** a man who was going to punch someone. (MAG[HarpersMag])
- (6.50) "You **look like** a man with something on his mind." (FIC[Analog])
- (6.51) She was **looking like** a woman who had no idea just how beautiful she was.
(FIC[Bk:YouBelievers])

(6.52) “You **look like** a woman in search of a cause.”
(FIC[LiteraryRev])

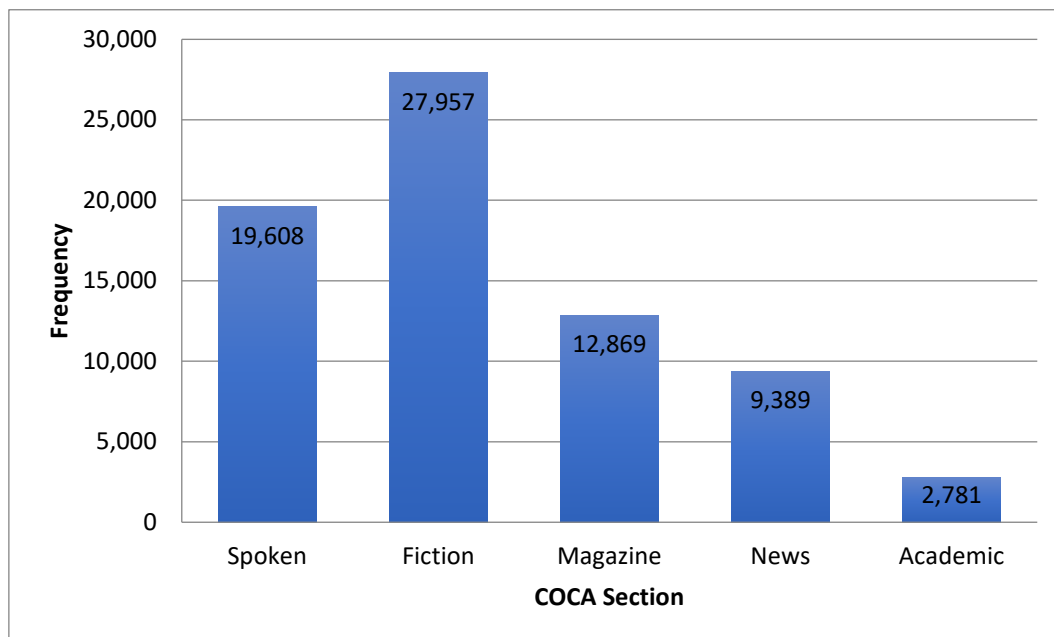
In the Event category, the collocates *accident*, and *robbery*, are mostly used for situational evaluation, but again, in many of these occurrences, the meaning could also be construed as visual description. These two collocates are often found in the *make NOUN VERB like NOUN* construction.

Finally, the irrealis collocates, *an angel*, *(a) monster/s*, and *(a) zombie/s*, by their very nature, would be associated with similes.

6.3.2: *LOOK like* Usage across Genres

Figure 7.1 presents the frequency distribution of *LOOK like* for each section of the COCA.

Figure 6.1: Frequency of *LOOK like* by COCA section

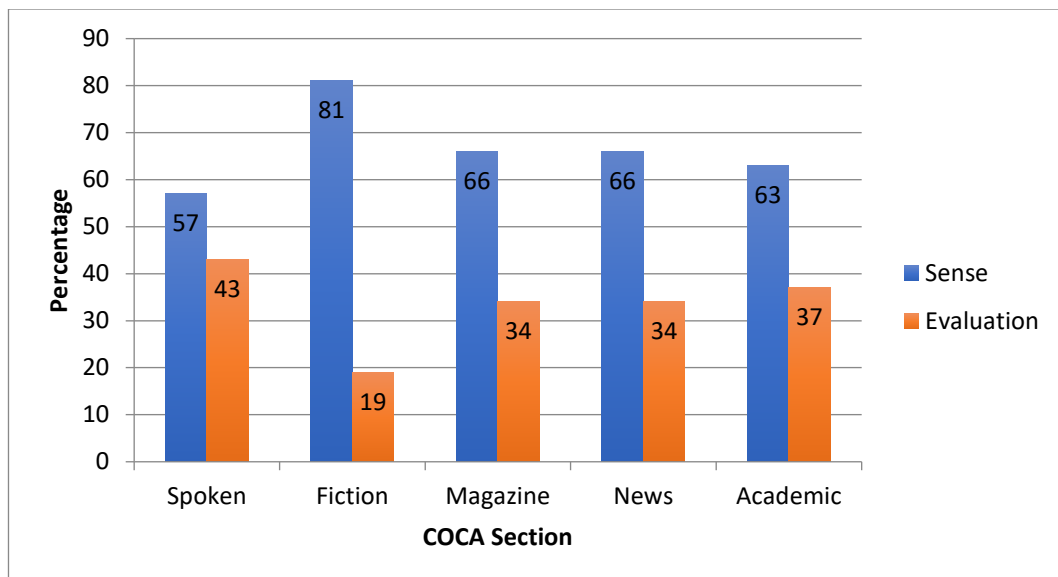


The data show that *LOOK like* is most common in the Fiction section, with nearly 28,000 occurrences, and the Spoken section, with just over 19,600 occurrences, and least common in the Academic section with less

than 2,800 occurrences. These findings closely coincide with those of Biber et. al (1999: 418) and their research with the Longman Spoken and Written English Corpus (LSWEC).

A closer examination of how usage of *LOOK like* varies across genres was carried out with samples of 100 concordance lines from each COCA section to investigate the frequency distribution of the main functions, or meanings, associated with *LOOK like*. These results are displayed in Figure 6.2 below.

Figure 6.2: Functions of *LOOK like* by COCA section



The data from the sample show that the description of visual sensory experience is the dominant usage of *LOOK like* across all five sections of the COCA. The data also shows that the ratio of visual sense description to situational evaluation is nearly identical for the Magazine, News, and Academic sections, averaging at 65 percent visual sense description and 35 percent situational evaluation. However, the two sections with the most frequent usage of *LOOK like*, Fiction and Spoken, have notably different usage ratios. The most divergent ratio between functions of *LOOK like* was

observed in the Fiction section, with visual sense description accounting for 81 percent of the sample, and only 19 percent being used for situational evaluation. This suggests that writers of fiction rely heavily on the *LOOK like* construction to describe the environments and characters they create. The Spoken section, on the other hand, showed the least divergence between functions of *LOOK like* with 57 percent of its usage being for visual sense description and 43 percent for situational evaluation, suggesting that situational evaluation is an integral component of spoken communication.

6.4: *FEEL Like*

The second most frequent PV phrase in the COCA, *FEEL like*, occurs 48,659 times and was also found to be highly phraseological, with collocates in six of Hanks' (2005) seven semantic groups. *FEEL like* is the most polysemous of the PVs with a total of four main functions identified in the data. In addition to the two core functions of sensory description and situational evaluation outlined previously, *FEEL like* is also used for, describing feelings and describing wants.

According to the corpus data, *FEEL like* is most often used for situational evaluation, accounting for around 50 percent of all occurrences. Using *FEEL like* in this way functions differently than *LOOK like*, however. It is often related to more personal issues, or it implies more of a personal opinion rather than an evaluation based on visual stimuli. Examples 6.53–6.56 show this form of evaluation.

- (6.53) We **feel like** consumer confidence is rising.
(NEWS[USAToday])
- (6.54) I **feel like** I failed. (NEWS[Denver])
- (6.55) I **feel like** that was kind of self-explanatory.
(NEWS[STLouis])
- (6.56) I want them to **feel like** it was a place they
could feel comfortable. (ACAD[MusicEduc])

Here again, there appears to be a fuzzy boundary between meaning and usage, in that some instances of *FEEL like* could be construed as being either situational evaluation on a personal level or the description of one's feelings. In many cases, however, the meaning of situational evaluation can be determined by glossing *feel like* with *think*, as in Examples 6.57–6.60.

- (6.57) We **think** consumer confidence is rising.
 (6.58) I **think** I failed.
 (6.59) I **think** that was kind of self-explanatory.
 (6.60) I want them to **think** it was a place they could
 feel comfortable.

The second most frequent use of *FEEL like* is for describing feelings, and this accounts for just over a third of its usage in the sample. Here again, the vehicles of the comparisons range from literal descriptions to very figurative and lexicalized similes.

- (6.61) I **feel like** I'm about to throw up,
 (SPOK[CNN_Presents])
 (6.62) We **feel like** kids in a candy shop.
 (NEWS[Atlanta])
 (6.63) Does your heart **feel like** it's in your throat?
 (SPOK[CBS])
 (6.64) She made you **feel like** a black sheep in your
 illustrious family. (MAG[Atlantic])

If we apply the same glossing test here, we can see that all but the first example will not work with *think*.

- (6.65) I **think** I'm about to throw up.
 (6.66) *We **think** kids in a candy shop.
 (6.67) *Does your heart **think** it's in your throat?
 (6.68) *She made you **think** a black sheep in your
 illustrious family.

This shows the distinction between the two uses, in which cases where the glossing of *think* does not work signifies that the intended meaning is to describe a feeling, but it also demonstrates the fuzzy boundary that

sometimes exists between them. At this boundary, “I feel like I’m about to throw up” can be understood as describing a feeling while at the same time functioning as situational evaluation.

The third most frequent use of *FEEL like* is concerned with describing wants or desires (like_{2.3}), accounting for approximately ten percent of the sample. This function is usually associated with the *FEEL like VERB-ing* pattern, where *FEEL like* is followed by a gerund.

- (6.69) Don’t **feel like** going out tonight? Curl up with
a DVD gift pack from Blockbuster (MAG[Redbook])
(6.70) She’d leave the welcome mat out for anytime he
felt like coming back. (FIC[Bk:MenOtherworld])
(6.71) And he **felt like** getting that tattoo.
(SPOK[CNN_Talkback])

There are also instances of *FEEL like NOUN* that are used for the function of describing wants or desires, although these occur much less frequently than *FEEL like VERB-ing*.

- (6.72) “Do you **feel like** a trip to Dallas to shop for
my wedding suit?” (Fic[GoodHousekeeping])
(6.73) And I **feel like** crab nachos.
(SPOK[NBC:TodayShow])
(6.74) They worked only when they **felt like** it and only
because they wanted to. (FIC[ContempFic])

Hanks’ (2005: 10) makes no mention of *FEEL like + gerund* taking the meaning of ‘want’ or ‘in the mood for’. Instead, he argues that this construction “exploits the same communicative function of evoking insight into private feelings by postulating a resemblance”. To support this claim, he provides the following example taken from the BNC:

- (6.75) Shiona **felt like striking him**. The man was
intolerable. (BNC:JXS[W_fict_prose])

He argues that like all similes, it is not appropriate to seek precision, and that one would not ask Shiona, “Where exactly did you want to strike

him, and how many times?” (Ibid:10). It could be argued, however, that this example is not actually a true simile, and this becomes evident when subjected to a reversibility test. Additionally, Example (6.72) from the COCA above, asking someone if they feel like a trip to Dallas to shop for a wedding suit is very specific and directly counters Hanks’ argument. This usage of *FEEL like* also seems to show that the *VERB like* construction is not always used to express similarity or to signal a simile.

The final, and least frequent, use of *FEEL like* is for the literal description of tactile sense, making up only six percent of its usage. Here also, the comparisons range from the literal to the figurative, as can be seen in the examples below:

- (6.76) . . . her fingers touched a hard object wrapped
in what **felt like** cloth . . .
(FIC[Bk:SilencedNovel])
- (6.77) The wheelchair **feels like** a huge heavy stroller.
(FIC[SouthwestRev])
- (6.78) It **felt like** his skin was burning away.
(FIC[BkSF:Nightshade])
- (6.79) After about a week, my **hair felt** like beef
jerky, so I asked him to wash it.
(MAG[Cosmopolitan])

6.4.1: *FEEL like* Characteristic Collocates

Looking at the characteristic collocates associated with *FEEL like* in Table 6.6, we can see that there are collocates that clearly link to its specific meanings, and most of these also fall within a specific meaning group.

The data in Table 6.6 show that *FEEL like* collocates with nouns in six of the seven semantic groups, with no statistically significant collocates occurring in the animal group. The most frequently occurring nouns in the *FEEL like NOUN* construction are words associated with human roles, and all of these nouns are associated with the function of describing feelings.

The majority of the Human Role collocates, with the exception of *pro/s*, are used to describe people negatively. This tendency for negative collocates is even stronger than it is for *LOOK like*, which also has a high proportion of negative collocates. Considering that the most frequent noun preceding *FEEL like* in the *NOUN FEEL like NOUN* pattern is the pronoun, *I*, this would seem to suggest that people often use *FEEL like* for self-deprecating purposes.

Table 6.6: Characteristic collocates of *FEEL like* in the COCA with MI scores above three, and frequencies over ten, sorted into the seven main semantic groups outlined by Hanks (2005)

Human Status	Human Role	Human Attribute	Animal	Artifact	Event	Irrealis
(a) family 32 (a) motherless child 11	(an) outsider 95 (a) stranger/s 46 (a) prisoner/s 35 (an) intruder/s 25 (a) fraud/s 23 a rookie 17 (a) foreigner/s 17 (a) traitor/s 12 (a) voyeur/s 12 (a) pro/s 11	a new person 17 a new man 15 second-class citizens 14		(a) home/s 180 ice 21 lead 20 sandpaper 18 rubber 12 (a) heel/s 12 (a) sponge/s 11 silk 10	(a) failure/s 90 work 21 spring 16 a betrayal 14 a burden 14 love 14 slap 14 the end of the world 12 a success 10	Cinderella 18

With the exception of *home/s*, all the other nouns found to significantly collocate with *FEEL like* in the Artifact category are associated with the function of describing tactile sensory experience. The Event category, in contrast, contains collocates from three of the four functions of *FEEL like*. The collocates *failure*, and *work*, for example are used for situational evaluation, while the collocates, *love*, is used to express feelings. Finally, the collocate *slap*, was found to be used for two different functions of the *FEEL like NOUN* construction, the description of tactile sensation, and describing feelings. Despite *slap* being a physical action, it is only used to describe a tactile sensation in three of its 14 occurrences in the COCA, all of which are used figuratively in descriptive similes in the Fiction section:

- (6.81) My friends were playing dodgeball, and every hollow bounce of the flaccid ball **felt like** a slap. FIC[Bk:DogWhoSavedMeNovel])
- (6.82) The night air **felt like** a slap, and not an unwelcome away. (FIC[SatEvenPost])
- (6.83) The air conditioner switches on, and its sudden exhalation **feels like** a chilly slap. (FIC[SouthwestRev])

The majority of the occurrences (11 out of 14) of *slap* in the *FEEL like NOUN* construction are associated with a lexicalized simile used to describe the personal feeling of being insulted or upset by someone's action or comment, and eight of these consist of the expression *FEEL like a slap in the face*. This expression was found in all five COCA sections with half of them coming from reported speech in the News section.

- (6.84) . . . and every passing hour felt like a rebuke, and every remark her mother had made **felt like** a slap, . . . (FIC[AntiochRev])
- (6.85) Shelley Blecha, a proponent for Nathan's Law in memory of her son who died in an unlicensed day care, said in an interview that the bill **felt like** a slap in the face. (NEWS[STLouis])
- (6.86) If a pat on the back in lieu of a raise can **feel like** a slap in the face, who dares complain? (MAG[Fortune])

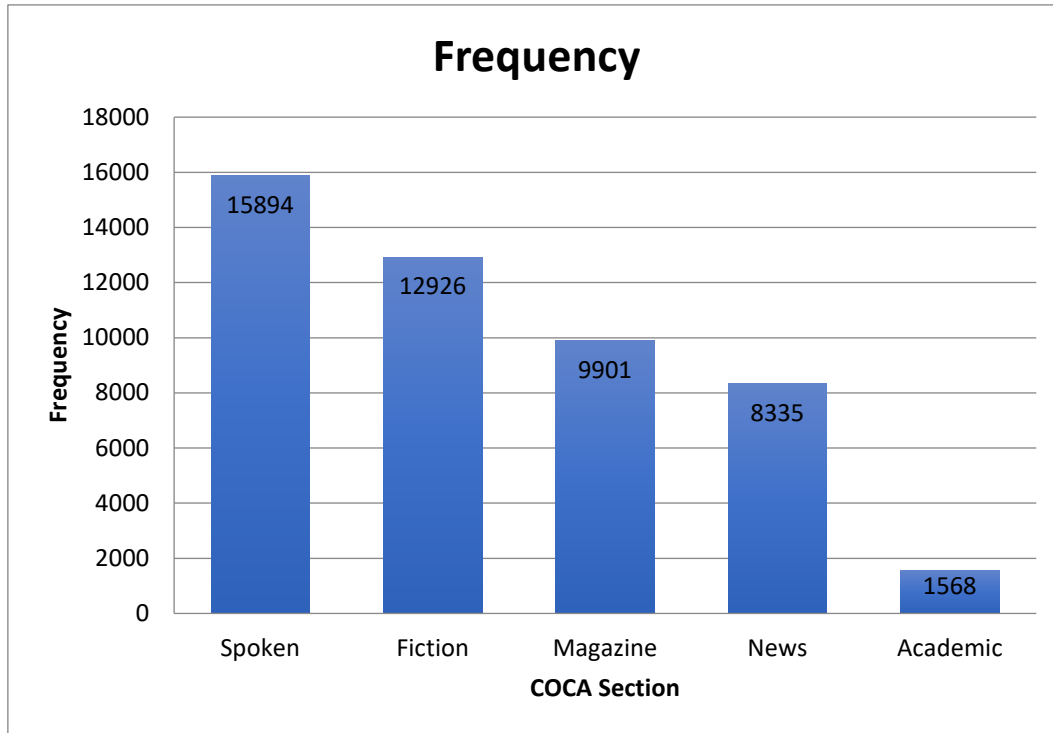
6.4.2: *FEEL like* Usage across Genres

Figure 6.3 shows the frequency data of the *FEEL like* construction for the five sections of the COCA.

These data show that *FEEL like* usage varies widely between COCA sections. It is most common in spoken English and fiction, and least common in academic writing. Interestingly, this frequency distribution differs from *FEEL* in isolation, which is most common in fiction in the COCA, and

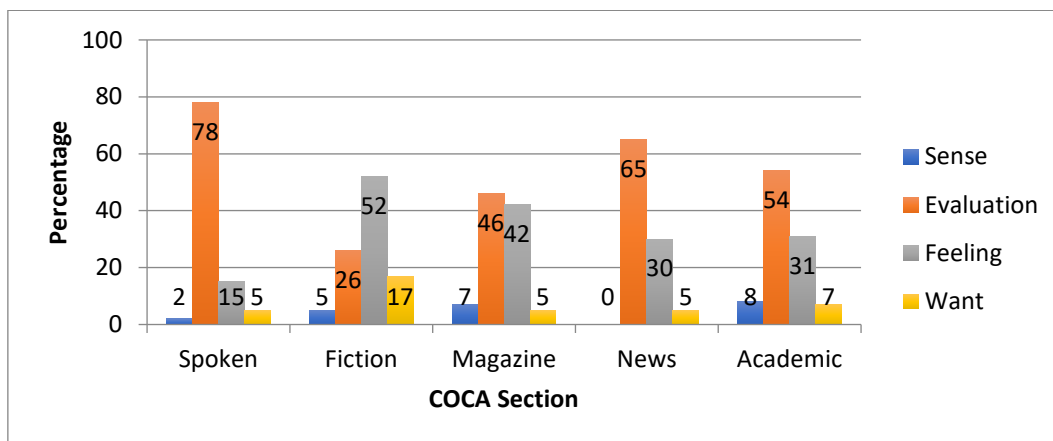
also in the Longman Spoken and Written English Corpus as reported by Biber et al. (1999: 368).

Figure 6.3: Frequency of *FEEL like* by COCA section



The frequency distribution for the various functions of *FEEL like* are shown in Figure 6.4 based on samples of 100 concordance lines from each section of the COCA.

Figure 6.4: Functions of *FEEL like* by COCA section



From these data, we can clearly see that the dominant function of *FEEL like* is situational evaluation in the Spoken, News, and Academic sections, with 78 percent, 65 percent, and 54 percent of the samples respectively. The most notable difference in usage occurs in the Fiction section where the most common usage of *FEEL like* is concerned with the description of feelings, accounting for over 50 percent of the sample, while this is the second most common function in the Magazine, News, and Academic sections with 42 percent, 30 percent, and 31 percent of the samples respectively. The Magazine section is the only section with a more balanced ratio of situational evaluation and describing feelings with 46 percent and 42 percent respectively. We can also see that describing wants is most common in fiction and that describing tactile sense using *FEEL like* is very uncommon in spoken English and news, with only 2 occurrences in the Spoken sample and zero occurrences in the News sample.

6.5: *SOUND Like*

The third most frequent *PV like* construction is *SOUND like*, with 23,175 tokens in the COCA. The two main functions of *SOUND like* include describing aural sense and situational evaluation, and as we have seen with the other PVs, the boundary between these two functions is at times rather fuzzy as well.

SOUND like was used for situational evaluation in approximately 53 percent of the time in a sample of 500 concordance lines. Examples of this usage from each of the five COCA sections are provided in the examples below, and it can be seen that *SOUND like* is used for situational evaluation in both *NOUN PV like NOUN* and *NOUN PV like CLAUSE* patterning.

(6.87) Is a \$1 Million Nest Egg Enough? It **sounds like**
a lot of money and it is. (MAG[Money])

- (6.88) That's beginning to **sound like** a third-world country to me. (NEWS[STLouis])
- (6.89) He **sounded like** they'd been talking for a good half hour. (FIC[TexasRev])
- (6.90) "Oh? What else does Miki say?" It **sounded like** Miki said quite a lot around here, actually. (ACAD[AmScholar])
- (6.91) It's **sounding like** it's going to be a game time decision. (SPOK[Fox_Susteren])

The examples above reflect the larger sample, which suggests that the speaker or writer usually bases situational evaluation with *SOUND like* on information that is heard or read. This fits with the findings from the *LOOK like* and *FEEL like* data samples where there are nuanced differences in the manner of situational evaluation based on the source of information as usually indicative by the main verb.

Next, *SOUND like* was used for the description of aural sense experiences in approximately 47 percent of the data sample. As with the other PVs discussed thus far, the comparisons range from literal to figurative, and both *NOUN PV like NOUN* and *NOUN PV like CLAUSE* patterning is used, although the majority of occurrences with *SOUND like* used for aural sense description are used in *NOUN PV like NOUN*.

- (6.76) Much of it **sounds like** it could be an early PiL record. (NEWS[Pittsburgh])
- (6.77) She said it **sounded like** techno music with a lot of bass, (NEWS[Denver])
- (6.78) to a predator the noise may **sound like** a mass of bees and have the effect of scaring it away. (MAG[NaturalHist])
- (6.79) One of the boars grunts, and it **sounds like** a dinosaur burping down a well. (MAG[MensHealth])
- (6.80) When you hit some railroad tracks, does it **sound like** you're driving inside a bongo drum? (MAG[PopMech])

6.5.1: *SOUND like* Characteristic Collocates

Table 6.7 below shows the characteristic collocates of *SOUND like* sorted into the seven main semantic groups outlined by Hanks (2005). From this data, we can see that *SOUND like* has significant characteristic collocates in only three of the seven groups, with the majority of these occurring in the artifact and event groups.

Table 6.7: Characteristic collocates of *SOUND like* in the COCA with MI scores above three, and frequencies over ten, sorted into the seven main semantic groups outlined by Hanks (2005)

Human Status	Human Role	Human Attribute	Animal	Artifact	Event	Irrrealis
	a republican 20 a democrat 10 a lawyer 10			(a) broken record/s 21 a cliché 20 an oxymoron 18 science 17 (a) firecracker/s 16 a train 13 a jet 12	a plan 45 a joke 33 (a) gunshot/s 29 thunder 24 music 24 a threat 23 an accusation 22 a question 19 good news 14 a nightmare 11 a challenge 10 a lie 10 a deal 10	

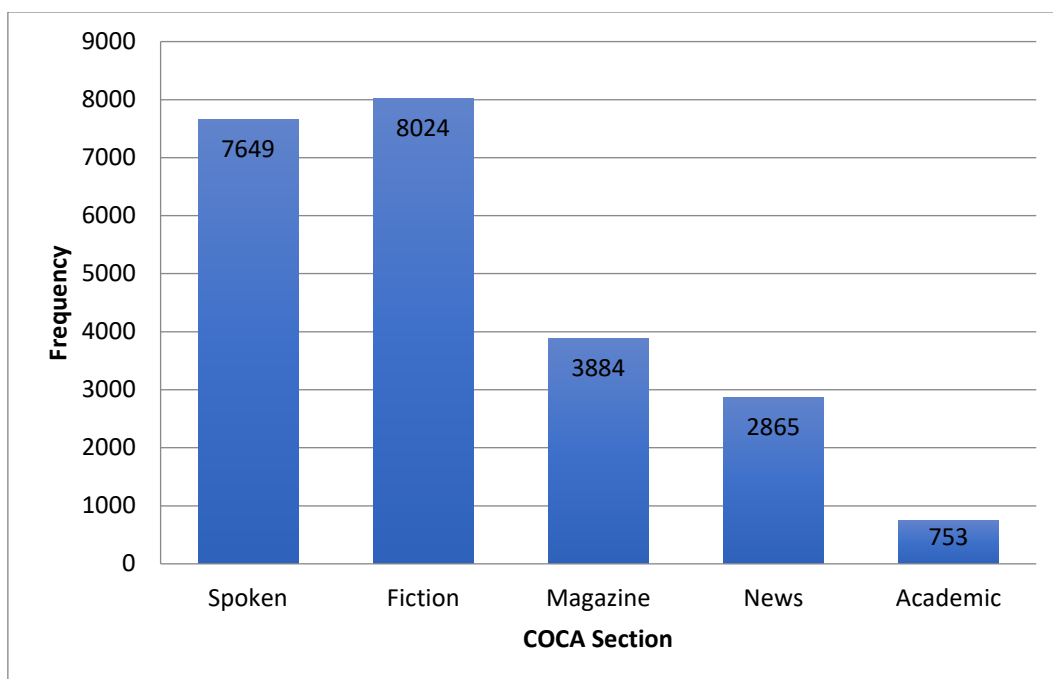
First, looking at the significant collocates in the event category, we can see that the majority of these nouns would be used for situational evaluation (e.g., *a plan*, *an accusation*, *good news*, and *a night mare*). Only three collocates in the event category, *gunshot/s*, *thunder*, and *music*, would typically be associated with the description of auditory sensory experience. Next, in the artifact category, there is a more balanced ration of collocates associated with both of the core meanings associated with *SOUND LIKE*, with four collocates used for situational evaluation (*broken record/s*, *a cliché*, and *an oxymoron* and *science*), and three associated with auditory sensory experience (*firecracker/s*, *train*, and *jet*). Finally, there are only three significant collocates in the human role category, and interestingly, all three of

them (*republican*, *democrat*, and *lawyer*) seem to be associated with negative situational evaluation in the realm of politics.

6.5.2: *SOUND like* Usage across Genres

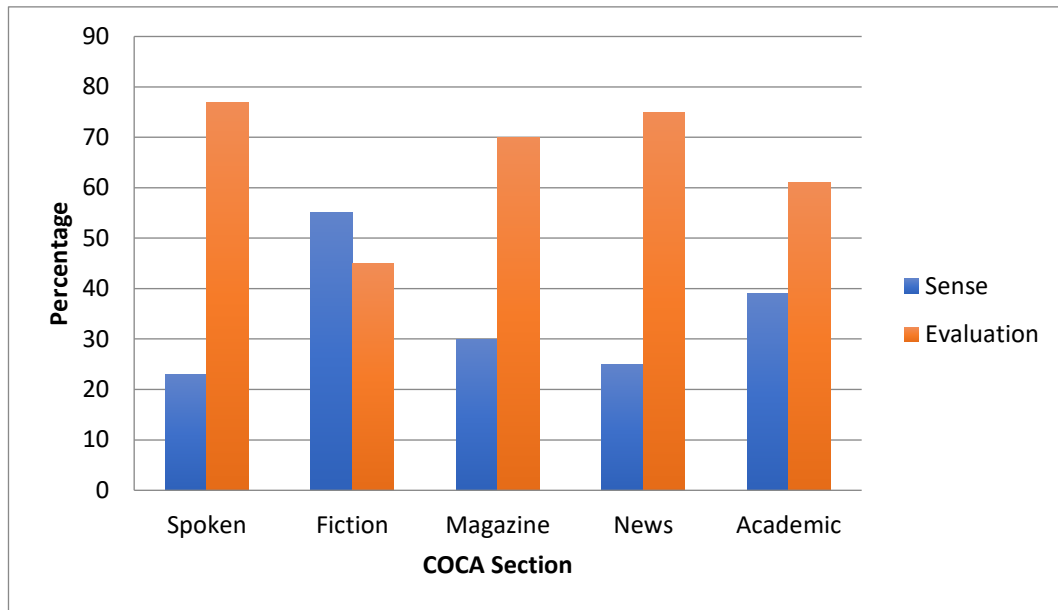
Figure 6.5 below shows the frequency of occurrence of *SOUND like* for each of the five COCA sections.

Figure 6.5: Frequency of *SOUND like* by COCA section



These data show that the *SOUND like* construction is most frequently used in the Fiction and Spoken sections, with 8024 and 7649 occurrences respectively. While these results are similar to both *LOOK like* and *FEEL like* in that all three *PV like* phrases are most common in these two sections, *SOUND like* is more evenly distributed between the two genres.

Next, in Figure 6.6, we can see how the two functions of *SOUND like* are used in the five sections of the COCA.

Figure 6.6: Functions of *SOUND like* by COCA section

These data show that the function of situational evaluation is the dominant function of *SOUND like* in four of the five COCA sections by a large margin. Once again, however, the Fiction section stands out with sensory description being more frequent than situational evaluation.

6.6: *SEEM Like*

With 20,771 tokens in the COCA, *SEEM like* is only slightly less frequent than *SOUND like*. While (as previously noted) it is not a true sense verb, it does share a proportionate amount of phraseology with *LOOK like*, *FEEL like*, and *SOUND like*. It also occurs much more frequently in the COCA than *SMELL like* or *TASTE like*.

Its primary function is for situational evaluation, and as discussed in 6.2.3, it shares collocates with all three of the polysemous PVs discussed above, all of which can only be used for situational evaluation (e.g., *bargain*, *betrayal*, *no-brainer*, and *eternity*).

- (6.81) He was elevated to almost a godlike status overnight, and politically it **seemed like** a good bargain. (SPOK[NPR_Morning])
- (6.82) To side with environmentalists, marine scientists, and other doomsayers against the wisdom of these oldtimers **seems like** a betrayal of our heritage. (MAG[NaturalHist])
- (6.83) It **seemed like** a no-brainer to Powell, whose business background made looking at the bottom line something of a reflex. (ACAD[THEJournal])
- (6.84) It **seemed like** an eternity ago, back when they were still in love. (FIC[Bk:TruthStainedLies])

This role in situational evaluation is further confirmed by looking at the characteristic collocates of *SEEM like* in Table 6.9. It is strongly associated with the event semantic group, in which the theme of time features prominently. Here, *SEEM like* is most closely related to *FEEL like*, sharing many collocates. These include *(an) hour/s*, *an eternity*, *(a) year/s*, and *a long time*. And, combined with its time-related characteristic collocates (such as *days*, *minutes*, and *weeks*), they account for nearly half of all the construction's collocates found in the event category.

Another semantic theme that features prominently in the event category of collocates is related to decision-making, including *a waste*, *a no-brainer*, *overkill*, *an afterthought*, *the right time*, *the best way*, *a bad idea*, and *a good opportunity*. Here, however, usage of *SEEM like* is most similar to that of *SOUND like*, with their shared decision-making collocates including *common sense*, *a great idea*, *a good thing*, *a good deal*, *a good way*, and *a good plan*.

6.6.1 *SEEM like* Characteristic Collocates

Table 6.9 shows the characteristic collocates associated with the *SEEM like* construction.

Table 6.9: Characteristic collocates of *SEEM like* in the COCA with MI scores above three, and frequencies over ten, sorted into the seven main semantic groups outlined by Hanks (2005)

Human Status	Human Role	Human Attribute	Animal	Artifact	Event	Irrealis
		(a) nice guy/s 40 (a) nice man/men 15		a good place 21 miles 11	a waste 40 a no-brainer 37 a big deal 33 a good time 29 overkill 22 ancient history 21 days 19 old times 16 an afterthought 15 the right time 14 the best way 14 minutes 13 a bad idea 12 a daunting task 11 ages 11 weeks 10 a good opportunity 10	a miracle 21 paradise 11

This data shows that while *SEEM like* has significant characteristic collocates in four of the seven semantic groups outlined by Hanks (2005), the vast majority are found in the event category. Furthermore, a closer look at the event-related collocates shows that almost all of them can be placed within one of two main semantic themes. First, the theme of decision-making features prominently in nearly half of the event collocates (e.g. a *no-brainer*, *the right time*, *a bad idea*). Second, four of the event collocates are closely associated with the theme of time passing (*days*, *minutes*, *ages*, *weeks*), or more specifically, the negative evaluation of time passing.

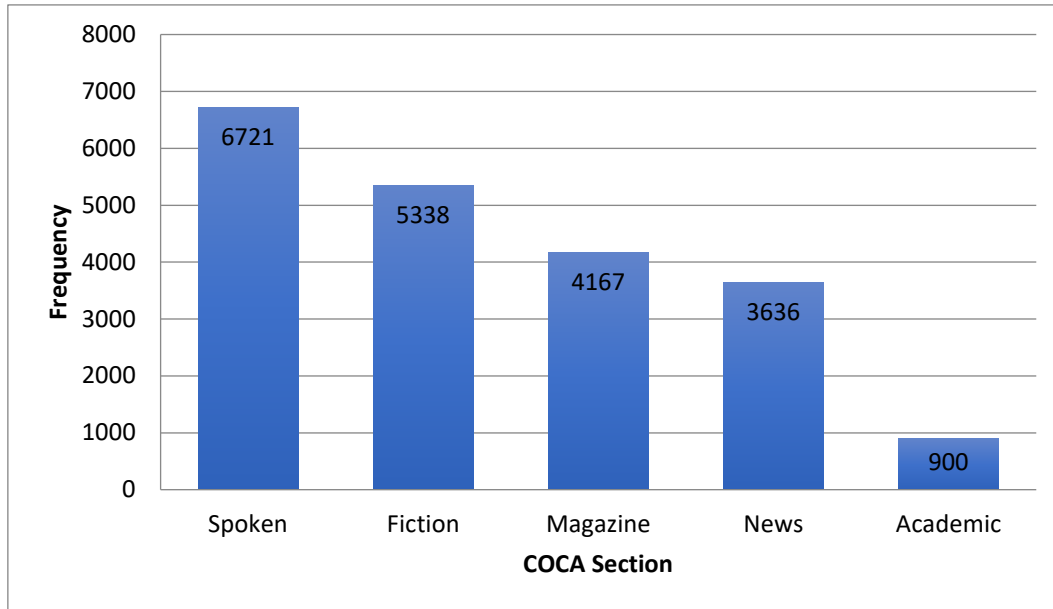
6.6.2: *SEEM like* Usage across Genres

The frequency distribution of *SEEM like* for the five COCA sections are shown in Figure 6.7.

The data show a recurring trend with all of the other PVs, that is, *SEEM like* is used the most in Spoken and fiction sections, and used the

least in the Academic section, with usage in the Magazine and News sections falling in the middle.

Figure 6.7: Frequency of *SEEM like* by COCA section



6.7: *SMELL Like*

Compared with the four preceding PV like phrases, *SMELL like* is significantly less frequent, with only 3,674 occurrences in the entire COCA and, as would be expected, it is much less phraseological as well. While *SMELL like* initially looks monosemous and used predominantly for the description of olfactory sensory experience (6.82–6.84), closer examination of the corpus data reveals instances of *SMELL like* used for situational evaluation. This usage was found to occur in approximately nine percent of the sample (6.85–6.86).

- (6.82) It's really appetizing. It **smells like** a grilled cheese sandwich. (SPOK[CBS_SunMorn])
- (6.83) His breath **smelled like** a mix of scotch and rancid butter. (FIC[FantasySciFi])
- (6.84) You go home **smelling like** a hamburger if you work at McDonald's. (NEWS[Chicago])

- (6.85) That **smells like** foreshadowing of both women making it to the title match.
(MAG[BleacherReport])
- (6.86) The Gregorian reform **smelled like** a Romish plot, and Ussher's contemporaries would be damned if they would accept it. (ACAD[NaturalHist])

6.7.1 *SMELL like* Characteristic Collocates

Table 6.14 below shows all the characteristic collocates of *SMELL like* with MI scores above three and frequencies of at least 10 in the COCA.

Table 6.10: Characteristic collocates of *SMELL like* in the COCA with MI scores above three, and frequencies over ten, sorted into the seven main semantic groups outlined by Hanks (2005)

Human Status	Human Role	Human Attribute	Animal	Artifact	Event	Irrealis
			(a) skunk 17	(a) rose/s 45 *teen spirit 38 sweat 18 soap 17 cinnamon 14 cigarettes 12 perfume 11 urine 10 gasoline 10 beer 10		

The data in the table show the characteristic collocates of *SMELL like* falling within only two of the seven semantic groups outlined by Hanks (2005). The vast majority of these collocates are artifacts and all but one of them are used for the description of olfactory sensory experience. When the word *roses* occurs in the *SMELL like NOUN* construction it is often part of the lexicalized simile, *smelling like roses*, and often part of the extended pattern *COME out smelling like roses*, used metaphorically for situational evaluation, to describe the conclusion of some sort of endeavor positively, and this is the case in five of the six times it occurs in the sample:

- (6.98) And Congress isn't **smelling like roses** either,
(SPOK[Fox:TheFive])

- (6.99) . . . bitches who could be dragged through shit
but somehow always came up **smelling like roses**.
(FIC[Bk:LowPressure])
- (6.100) Will the two heroes come out **smelling like roses**—like Paul Newman and Robert
Redford . . . (ACAD[literaryRev])
- (6.101) The President got away, in a sense, with being
in Maine, and we come out **smelling like roses**.
(SPOK[CNN_King])

Also of note is the highly frequent collocate *teen spirit* in the Event category. This is always referring to the famous song by the group Nirvana, and although it is only a song name, the phrase is based on situational evaluation.

Next, there is one non-event related collocate found in the animal category, with 17 occurrences of the lexicalized simile *SMELL like a skunk*.

- (6.102) He **smelled like** a skunk. He needed a bath, a
place to rest. (Fic[FantasySciFi])

6.7.2: *SMELL like* across Genres

The frequency data for *SMELL like* for each of the five COCA sections is displayed in Figure 6.8.

These data show a marked difference between *SMELL like* and the other PVs in that the vast majority of its usage occurs in the Fiction section of the COCA. While Fiction is one of the top genres for all of the PVs discussed thus far, here we can see that the frequency of use is much lower in the remaining four genres when compared to the other PVs. This adds further weight to the previously noted findings showing that fiction makes use of the *PV like* construction for sensory description notably more than the other genres.

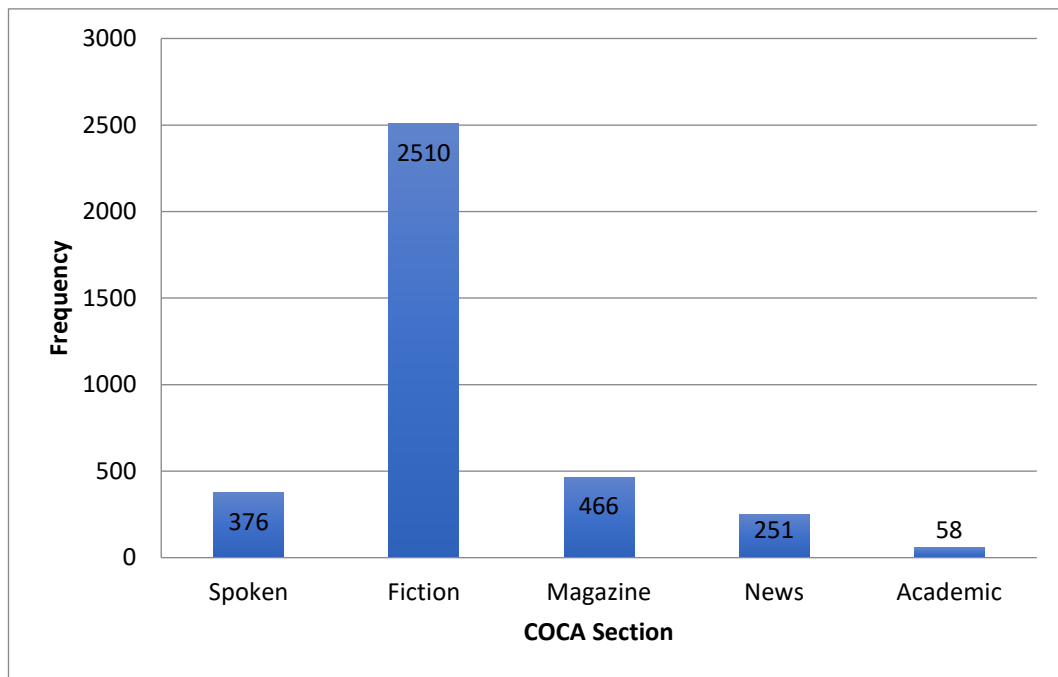
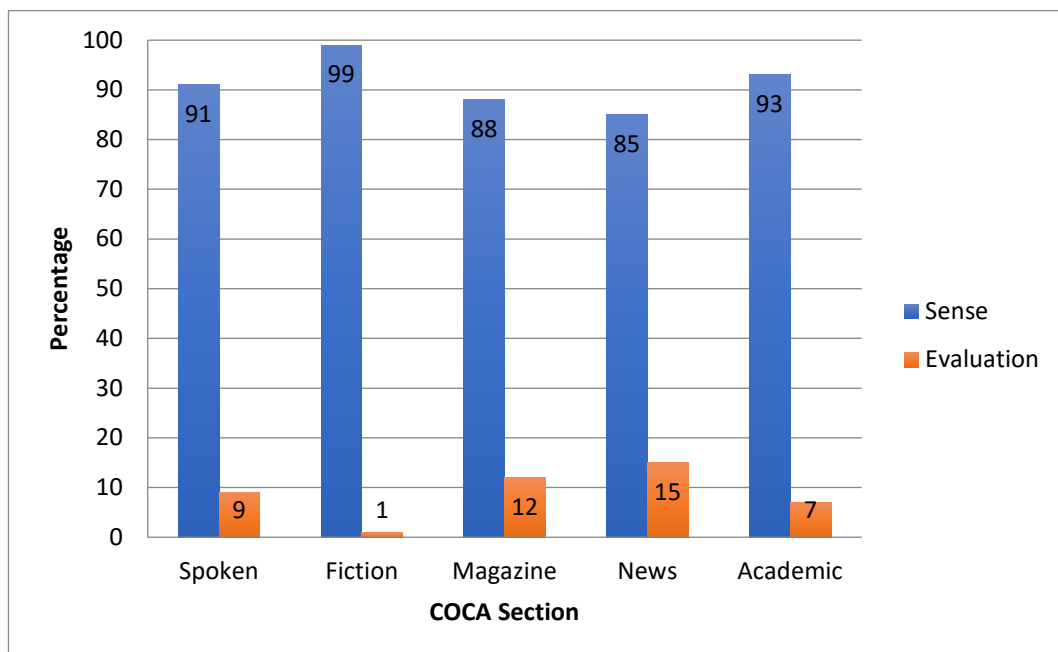
Figure 6.8: Frequency of *SMELL like* by COCA section

Figure 6.9 shows the proportion of usage for the functions of *SMELL like* in each of the five COCA sections.

Figure 6.9: Function of *SMELL like* by COCA section

These data show similar results for all five sections. As previously stated, *SMELL like* is used predominantly for the description of sensory experience across all genres. While rather uncommon, *SMELL like* is used for situational evaluation most often in the news and magazine sections with 15 percent and 12 percent of the samples respectively, and it is almost never used for this function in fiction, with only one occurrence in the sample.

6.8: *TASTE Like*

Finally, we come to the least frequent and least phraseological of the *PV like* phrases, *TASTE like*, which occurs only 1,913 times in the COCA, and like *SMELL like*, *TASTE like* is mostly monosemous and usually used for the description of its respective sense. It was found to be used for describing sense experience in 98.4 percent of the sample (Examples 6.103 and 6.104) and for situational evaluation in only 1.6 percent of the sample (Example 6.105).

(6.103) My grilled cheese, I like a grilled cheese that
tastes like a calzone, so I take parmesano
 164eggiano, fresh ricotta cheese.

(SPOK[ABC_GMA])

(6.104) Moose **tastes like** a good filet mignon. It's not
 fatty. (NEWS[Atlanta])

(6.105) The brandy from last week **tasted like** horse
 piss. (FIC[Bk:SFThreeMusketeers])

(6.106) Nearly every sentence, every phrase governed by
 a world-weary resignation that **tastes like**
 doubt. (FIC[SouthwestRev])

6.8.1: *TASTE like* Characteristic Collocates

Table 6.11 shows the statistically significant characteristic collocates of the *TASTE like* construction.

Table 6.10: Characteristic collocates of *TASTE like* in the COCA with MI scores above three, and frequencies over ten, sorted into the seven main semantic groups outlined by Hanks (2005)

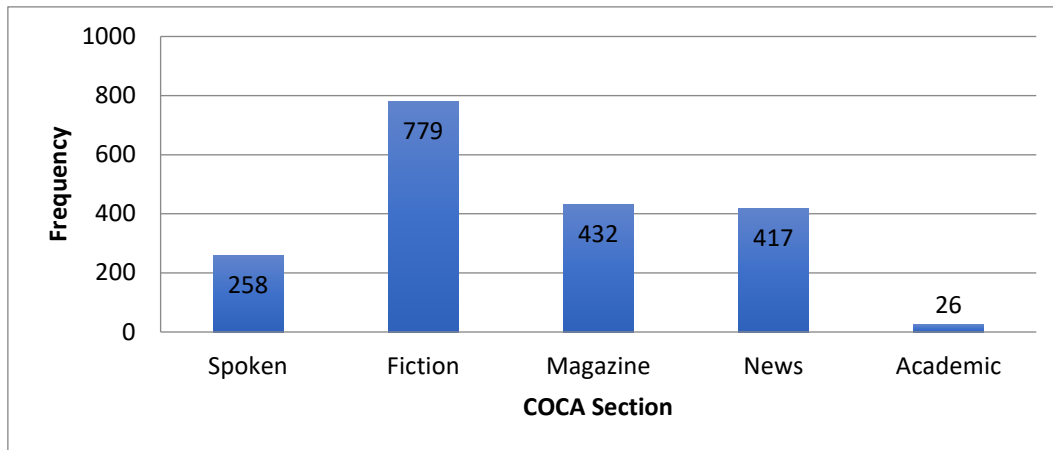
Human Status	Human Role	Human Attribute	Animal	Artifact	Event	Irrealis
			chicken 42	cardboard 23 dirt 12		

These data show us that, similar to *SMELL like*, *TASTE like* is not nearly as phraseological as the other *PV like* phrases, with only three characteristic collocates. Here, too, many nouns that co-occur with *TASTE like* occur fewer than ten times, and there are many hapaxes. This suggests that *TASTE like* is also used more in line with the open-choice principle, and that it requires a certain level of language creativity from speakers. Also similar to *SMELL like*, the majority of the characteristic collocates that occur with *TASTE like* are artifacts, and both of them (*cardboard* and *dirt*) are used in similes to describe something as tasting very bad. The third characteristic collocate, *chicken*, was placed in the animal category (although it could also be argued to belong in the artifact category as it usually represents the meat of a chicken) and is generally used in a more literal and neutral way.

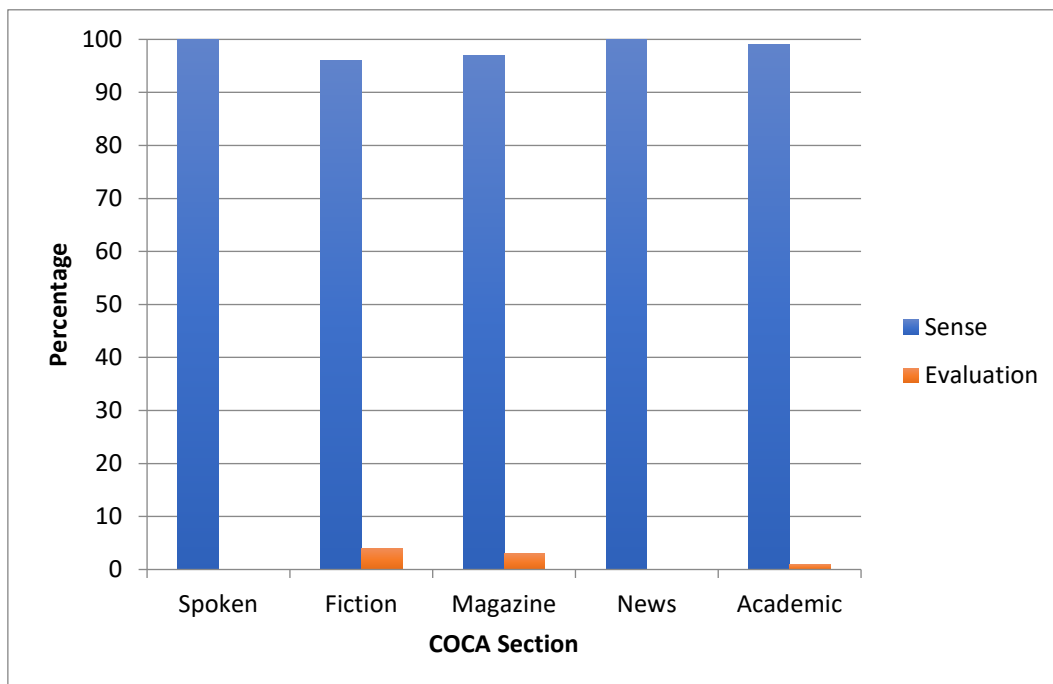
6.8.2: *TASTE like* Usage across Genres

Figure 6.10 below shows the frequency data for *TASTE like* across the five genres of the COCA.

These data show a very similar frequency distribution between *SMELL like* and *TASTE like* with the majority of occurrences in the Fiction section, and the least amount of usage in the Academic section.

Figure 6.11: Frequency of *TASTE like* by COCA section

Next, Figure 6.11 displays the percentage of usage for the functions of *TASTE like* across sections of the COCA.

Figure 6.12: Function of *TASTE like* by COCA section

These data also show similar results to *SMELL like* although, *TASTE like* is used for situational evaluation even less than *SMELL like*, with zero occurrences in the Spoken and News sections.

6.9: Conclusion

This chapter has looked at the *PV like* construction, focusing on the main functions and phraseological behaviours of the six perception verbs most frequently used with prepositional *like* (*LOOK*, *FEEL*, *SEEM*, *SOUND*, *SMELL*, and *TASTE*). First it should be noted that the frequency data for the *PV like* constructions generally confirms the work of Viberg (1983), Sweetser (1990), and Winter (2019), and this suggests there is a hierarchy of sensory description. Visual is by far the most frequently used, and both smell and taste are the least frequently used, with sound and touch in the middle. However, as I have shown, for a large portion of their usage when paired with *like*, the majority of the sense verbs do not behave as true sense verbs at all. Instead, the most frequent perception verbs—*LOOK*, *FEEL*, and *SOUND*—appear to be polysemous, with the two core meanings of situational evaluation and the description of sensory experience. Moreover, the verbal phrases formed with all three of these PVs share the related function of situational evaluation, showing nuanced differences between the verbs. In the cases of *FEEL like* and *SOUND like*, this figurative usage occurs more frequently than the literal description of sense. Additionally, *FEEL like* was found to be the most polysemous, with the additional functions of describing one's feelings and describing what one wants with the *FEEL like VERB-ing* construction. Only the much less frequent *SMELL like* and *TASTE like* were found to mostly behave as true sense verbs, although these PVs are also occasionally used for situational evaluation. Finally, the data shows that the PVs are used for sensory description in fiction more often than in the other genres in the COCA.

7

Results

Functions of *Like* When Premodified with *Just*

7.1: The Problem at the Single-Word Level

In this chapter, by means of a corpus-based analysis of the co-occurring words *just* and *like*, I aim to demonstrate the problem with traditional lexical and grammatical classification at the single-word level. I will show how single-word entries in traditional dictionaries are often not helpful when dealing with multi-word lexical constructions, and how traditional grammars are often too simplistic. This means that such references (and language teaching based on them) are only partially helpful and, in some cases, misleading for language learners. The analysis will involve comparing corpus data with three different linguistic reference resources. The first two—a comprehensive reference grammar (Downing and Locke 2006 (D&L)) and an online dictionary of American English (*Merriam-Webster* (M-W), accessed September 26, 2020)—represent the traditional, generative, slot-and-filler view of language, which Taylor (2012) aptly termed the dictionary-and-grammar model. I will also refer to a third resource, the corpus-based *Collins COBUILD Learner's Dictionary* (2003; CCLD). This way, I aim to determine if a more modern, corpus-based dictionary, that incorporates patterns and phrases, is able to account for the various lexical patterns and phrases I find associated with the *just like* construction.

Starting in Section 7.2, I will review the lexical representations of *just* and *like* as found separately in the M-W dictionary. I will then, in Section 7.3, review the traditional grammatical classification of premodification and provide frequency data for the various premodifiers found to occur with *like* in the Corpus of Contemporary American English (COCA). I will also review the rationale for choosing to focus solely on the *just like* construction. Next, in section 7.4, I will outline and discuss the most frequent patterns and phrases involving the *just like* construction as they conform to the traditional grammatical classification presented in Section 7.2. In section 7.5, I identify and analyze two *just like* constructions that cannot be explained using the dictionary-and-grammar model. And, finally, in Section 7.6, I will refer to the *CCLD* to determine whether or not the patterns and phrases identified in Sections 7.5 and 7.6 are included in this more modern, corpus-based dictionary. The layout of this chapter shows that I will first refer to a dictionary. Then, as needed, I will check in a reference grammar. This arrangement of tasks is meant to emulate the way in which teachers and learners would approach an item such as *just like* in a more traditional approach.

7.2: Lexical Representation of *Just* and *Like*

To demonstrate the issues that a student of English may confront when first noticing an unknown multi-word construction such as *just like*, we will first look at a typical example of *just like* usage. Example 7.1 below is taken from the first concordance line in a KWIC (keyword in context) search in the 612-million-word COCA corpus.

(7.1) His front and back legs moved in unison, **just**
like a bear's. (FIC[SouthernRev])

An English learner who felt the need to understand the meaning of *just like*, after reading this sentence would likely start (as most learners do) at the

lowest, single-word level. That is, they would look up each of the target words in a dictionary. Considering that the data in this study is taken from the COCA, we will first refer to a dictionary of American English. Both words are polysemous according to their respective M-W definitions for language learners, which are summarized below.

7.2.1 *Just*

The online M-W dictionary lists *just* as being in the top 10 percent of looked-up words and consists of three entries, summarized in Table 7.1.

Table 7.1: Summary of definitions for *just* in *Merriam-Webster* dictionary (online version; M-W)

Entry	Definition
<i>just</i>₁ (adjective) Sense	
1a	REASONABLE— <i>had just reason to believe he was in danger</i>
1b	PROPER— <i>just proportions</i>
1c	ARCHAIC— <i>faithful to an original</i>
2a (1)	RIGHTEOUS— <i>a just war</i>
(2)	DESERVED— <i>a just punishment</i>
2b	LAWFUL— <i>just title to an estate</i>
<i>just</i>₂ (adverb) Sense	
1a	EXACTLY, PRECISELY— <i>just right</i>
1b	very recently— <i>the bell just ran</i>
2a	by a very small margin: BARELY— <i>just too late</i>
2b	IMMEDIATELY, DIRECTLY— <i>just west of here</i>
3a	ONLY, SIMPLY— <i>just last year—just be yourself</i>
3b	QUITE, VERY— <i>just wonderful</i>
4	PERHAPS, POSSIBLY— <i>it just might work</i>
	<i>just about</i> : ALMOST— <i>the work is just about done</i>
<i>just</i>₃	archaic variant of JOUST

As can be seen in Table 7.1, there are a total of 15 senses of *just* in three separate entries. It may be likely that many learners would be able to deduce the meaning of *just* that they are looking for—as found under the *just*₂

adverb entry. However, there are still eight separate possibilities listed in four senses and one phrase (*just about*), and all this information could certainly prove problematic for the learner.

Scrolling down the page, there is a short list of seven phrases related to *just*. Only one of these phrases is concerned with *just like*: *I'd (just) like to see him try*. Clicking on this phrase brings the user to a separate page, on which the phrase is labeled an idiom with the following meaning:

—used to say that another person has no chance of succeeding at something // “He thinks he can beat you.” “*I'd (just) like to see him try!*”

In this phrase, *like* is clearly being used as a verb and is no part of the meaning of *just like* that we are trying to understand from Example 7.1.

Scrolling further down the page, the M-W also provides a separate section of definitions of *just* for English language learners. While the learner definitions are simplified, there are still six possible meanings listed for *just*. And there is still no clear explanation of what *just like* means.

7.2.2: *Like*

Not being able to understand the meaning of *just like* in Example 7.1, the learner might move on the second word, *like*. According to M-W, *like* is in the top one percent of looked-up words, and this very high frequency is reflected in the multitude of possible meanings found on the webpage. There are a total of 42 senses of *like* listed, under nine separate entries. This is even more complicated than the six entries found in the OED (outlined in Chapter 1, Section 1.4). The nine entries are summarized in Table 7.2.

Table 7.2: Summary of definitions for *like* in M-W

Dictionary Entry	Definition
<i>like</i> ₁ (verb)	enjoy, regard, want, to do well in, approve, choose/prefer
<i>like</i> ₂ (noun)	liking/preference, something that one likes
<i>like</i> ₃ (adjective)	the same or nearly the same, likely
<i>like</i> ₄ (preposition)	having the characteristics of / similar to, typical of, comparable to / approximating, in the manner of / similarly to, as though there would be, such as, used for intensive or ironic phrases: <i>fought like hell</i>
<i>like</i> ₅ (noun)	one that is similar/counterpart/equal, kind, one of many that are similar to each other
<i>like</i> ₆ (adverb)	equally, likely/probably, to some extent / rather / altogether, interjectional use, nearly/approximately
<i>like</i> ₇ (conjunction)	as if, in the same way that / as, in the way or manner that, discourse marker, such as
<i>like</i> ₈ (auxiliary verb)	came near / was near
<i>-like</i> ₉ (adjective combining form)	resembling or characteristic of

Considering the very large number of possible meanings of *like*, as shown in Table 7.2, it seems likely that identifying the correct meaning found in Example 7.1 could prove problematic, if not overwhelming, for the language learner. The M-W dictionary lists only four phrases related to *like*, none of which revealed much about the *just like* construction. And the definitions for language learners, although simplified, still consisted of 12 senses in five main entries.

7.3: Grammatical Classification of Premodifiers

Being unable to determine the meaning, or meanings, of the *just like* construction in the dictionary, the language learner might refer to a grammar. If the learner was able to identify *like* as a preposition, then they might try to understand the function of *just* as a premodifier of *like*.

In their comprehensive reference grammar, D&L note that, in addition to nouns, adjectives, and adverbs, prepositions can also be modified. They list several forms of premodification that are generally used with prepositions, from grading and intensification to directional forms and quantification, and from descriptive/attitudinal to focusing/reinforcement. The

corpus data show that *like* is used with all of these forms of premodification, except with directional premodification (*down, up, over, under*, etc.). Grading premodifiers, which involve the comparative and superlative forms, will be covered in detail in Chapter 8, and will therefore not be included in the following discussion. The remaining forms of premodification, used with prepositional *like*, are outlined below.

In D&L, the authors describe intensification across three separate sections: in Module 52.2, “Intensifying the Attribute” (488–493), in Module 56.2, “Intensifying the Adverbial Meaning” (516–517), and in Module 57.5, “Realisations of the Modifier Element” (538–539). While only Module 57.5 specifically addresses the premodification of prepositions, examples of premodifiers from all three sections were found to work with *like* in the COCA (see Subsection 7.2.1 below).

In these sections, the authors describe intensification as a form of grading, and they list three degrees of intensifying premodifiers: high intensification, medium intensification, and attenuation. High intensification is expressed by a wide range of adverbs and adjectives and, less frequently, with nouns. On the other hand, according to this grammar reference, medium intensification is limited to just four adverbs: *quite, pretty, rather*, and *fairly* (516). Attenuation, which they refer to as “a slight degree of the quality or its entire absence” (490) can be expressed with a slightly wider range of lexical items: *slightly, a little, a bit, kind of, sort of*, and *somewhat*. Moreover, it can also be expressed by negating a high level of intensification: examples include *not very, not quite, not entirely*, and *not particularly*. Additionally, D&L describe another group of attenuating premodifiers, ones that express “a minimal degree of attribution and often imply a certain degree of the opposite quality . . . they express absence or denial of the quality named” (491)

(*hardly, barely, scarcely, none too, not at all*). The authors point out that these three degrees of intensification should be thought of as a cline and not as a scale with fixed points, because they function through the use of lexical items rather than structural variation (488). They also point out that varied stress and intonation patterns can be used by speakers to adjust and reinforce the intended degree of intensification.

According to D&L (2006) there are two types of quantifying premodifiers. These are used in what the authors call exact quantification and non-exact quantification. Exact quantification is used to answer questions such as *How old is she?* And *How high is Everest?* Non-exact quantification, on the other hand, is used to express non-measurable quantification. Examples provided by D&L for this type of quantification are expressed through the use of determinatives (*the, that, this, any, all, little, and no*): *The trip wasn't **that** interesting after all; We need a box **this** big* (491). While this exact / non-exact distinction is found in their discussion of adjectival and adverbial groups, D&L make no mention of the two forms of quantifying premodifiers in their discussion of quantifying premodifiers. Instead, they simply provide a list of quantifying premodifiers used with prepositions that mostly consist of non-exact quantifiers (*a long time, not that much, miles, way back, light years, streets, nearly and almost*), with one exact quantifier (*two hours*) in the list (539).

According to D&L, there are two types of descriptive/attitudinal premodifiers used with adjectival and adverbial groups: qualitative and relational. Qualitative premodification can be expressed with *-ly* adverbs (e.g., *strangely, deathly*), adjectives (e.g., *light brown, deep red*), and nouns (e.g., *pitch black, paper-thin*), while relational premodification “indicates the sense in which the adjective is to be understood” (492). Relational premodifiers

can be expressed with *-ly* adverbs (e.g., *socially acceptable*, *economically difficult*), and with nouns (e.g., *girl-crazy*, *duty-free*). When discussing descriptive/attitudinal premodifiers used with prepositions, D&L offer only a brief list of *-ly* adverbs commonly used with prepositions, including *surprisingly*, *hopelessly*, *dangerously*, and *unexpectedly* (539). It is worth noting that, while prepositions are not supposed to be descriptively premodified with adjectives and nouns, *like* is somewhat unique. This is because it can be used in the *-like* adjective form to serve a similar function with both nouns (*girl-like*), and occasionally adjectives as well (*crazy-like*).

The final types of premodifier used with prepositions, as listed in D&L's grammar, are focusing/reinforcing premodifiers. They provide a list of example premodifiers in this category, including *precisely*, *mainly*, *just*, *principally*, *chiefly*, *merely*, and *only*.

For a language learner, D&L's categorization of premodifiers could prove to be somewhat confusing, as the topic of modification is spread across several sections. Specifically looking up the use of *just* could be even more problematic, as it is shown as an example in two types of premodification. They list *just* as an intensifier, along with *very* (as in *very soon*), for adverbs (*just then*; 516). They also list *just* as a focusing/reinforcing premodifier, as in *You may say that just for the sake of arguing* or *We arrived just before midnight* (539).

7.3.1: Premodifiers Used with *Like* in the COCA

In Table 7.3, I provide the frequency data for all the premodifiers found to occur with *like* in the COCA. They are sorted into the categories outlined by D&L, with the exception of the comparative/superlative premodifiers (these will be covered in Chapter 8).

Table 7.3: Raw frequency of premodifiers found to occur with *like* in the Corpus of Contemporary American English (COCA), sorted into Downing and Locke's (2006) grammatical categories. Premodifiers with frequencies lower than ten are not included. Eleven items (+) are not identified in D&L; *just* occurs in two (*) categories.¹ See 7.2.2.

Intensification High	Medium	Attenuation	Grading	Descriptive/ Attitudinal	Focusing/ Reinforcing	Quantifying
just** (20,384)	quite (1,151)	almost (4,747)	more (13,349)	actually (439)	just* (7,233)	almost (4,747)
exactly+ (1,836)	rather (732)	kind of (4,622)	less (850)	particularly (335)	only (213) pre-	always+ (986)
much (6,357)	pretty much+ (293)	sort of (2,548)	most (476)	probably (307)	cisely (40)	never+ (264) some-
very much (1,852)	more or less+ (81)	a bit (1,387)	more and more (318)	especially (257)	merely (16)	times+ (170)
so much (683)		a little (1,335)	the most (62)	suspiciously (206)		often+ (96)
too much (898)		a little bit+ (745)	at least+ (51)	remarkably (179)		occasionally+ (12)
really (4,282)		somewhat (415)	least (35)	vaguely (99)		
a lot (2,308)		not at all (246) re-	less and less (27)	strangely (39)		
completely (24)		motely (147)	the least (19)	surprisingly (35)		
absolutely (19)		slightly (41)	in the least (10)			
		hardly (12)				

As can be seen in Table 7.3, *like* is very frequently premodified across all of the categories described by D&L, with the exception of directional premodification. However, the frequency data in the table is somewhat misleading because in many cases the same premodifiers can be used with both prepositional *like* and with verbal *like*. Further complicating this, it turned out that many of the premodifiers are often shared amongst the three main meaning groups for *like*, necessitating a level of disambiguation that would reach beyond collocation and into the realm of grammatical patterning.

Intensification is the most frequent form of premodification, and premodifiers from all three intensity levels collocate with *like* in the COCA. It was found that many of the intensifiers listed in D&L's grammatical framework are used with *like*—or, at least, variations of these intensifiers are used. For example, with medium intensification, there are no instances of *pretty like* in the COCA where *pretty* is used as an intensifier (see the simile in

1. The two figures for *just* are estimates based on the findings in 7.3.2—that, in a data sample, intensifying *just* was almost always prepositional, and focusing *just* was always verbal. Based on this finding, I carried out two separate corpus queries, one for *just like* with *like* isolated as a preposition, and the other for *just like* with *like* isolated as a verb.

Example 7.2). However, there are numerous instances of *pretty much like*, which functions with a medium degree of intensification (see Example 7.3).

(7.2) Soline's apartment is **pretty like** a magazine, but
it's lonely in the evenings. (FIC[Bk:NearerHome])

(7.3) It sounds **pretty much like** a no-brainer.
(SPOK[NBC_Dateline])

In Example 7.2 above, we can see that *pretty* is used as an adjective in an adverbial simile to describe an apartment. Meanwhile in the simile shown in Example 7.3, *pretty much* modifies the *SOUND like* construction to show a medium degree of intensification. Additionally, many other premodifiers not mentioned by D&L were found to be used with *like*, as marked with a (+) in Table 7.3. And this comes as no surprise, since it would be nearly impossible to list in a reference grammar every premodifier used with every preposition.

Although *like*, as shown in Table 7.3, is very frequently used with numerous premodifiers and across several categories of premodifiers in the COCA, two premodifiers stand out as forming, by far, the most commonly used *premodifier like* constructions. There are 28,401 (46.4. per million) occurrences of *just like* in the COCA, a construction used as both an intensifying premodifier and a focusing premodifier; and there are 13,349 (21.8 per million) occurrences of *more like*. In keeping with the principles of the lexical syllabus (Sinclair and Renouf 1988) for further, in-depth analyses, I focused on the most frequent patterns of usage for our highly frequent target word *like*. Accordingly, only these two highly frequent *like* constructions were selected. The *more like* construction will be covered in Chapter 8's broader analysis of comparatives and superlatives used with *like*, and the remainder of the current chapter will focus specifically on the functions of the *just like* construction.

7.3.2: Frequency Distribution of *Just Like* Functions

As shown in Table 7.3, the *just like* construction is the most frequent *pre-modifier + like* construction in the COCA, with an initial corpus query returning 28,401 total occurrences. However, further analysis, prompted by D&L's placement of *just* in two premodifier categories, shows that *just* is used with all three of the main *like* functions. The results of this analysis, which involved sorting and categorizing a 500-concordance-line sample of *just like*, are shown in Table 7.4.

Table 7.4: Frequency distribution of *just like* across the three main functions of *like* in a 500-concordance-line sample taken from the COCA

<i>like</i> function	Frequency	Percentage
<i>like</i> ₁ similarity	452	90.4%
<i>like</i> ₂ fond/enjoy	29	5.8%
<i>like</i> ₃ pragmatic	19	3.8%

The data in Table 7.4 show that the vast majority of *just like* occurrences in the COCA (90.4%) are associated with prepositional *like*₁, used to show similarity (see Examples 7.4 and 7.5). And in all but one of these occurrences (Example 7.6), *just* acts as an intensifier.

- (7.4) At the tender age of 7, Chris wanted to be **just like** them. (News[AssocPress])
- (7.5) It would keep going until dawn, **just like** the six previous nights (FIC[Bk:Illicit])
- (7.6) But probably once I was, like, an adult, then you sort of realize like—oh, this isn't **just like** an attitude. This is actual—there may be something clinically happening here. (SPOK[NPR_FreshAir])

In Example 7.6, *just* can be glossed with the word *exactly*, and its effect is intensification of the degree of similarity being expressed.

A much smaller percentage of the *just like* occurrences in the sample (5.8%) involve verbal *like*₂ usage and, in all of these instances, *just* acts as a focusing premodifier (see Examples 7.7 and 7.8).

(7.7) But she didn't **just like** to read to herself—she liked to read aloud to her students.
(FIC[AdultLearning])

(7.8) I mean, I'd **just like** to speak with a lawyer.
(SPOK[ABC:20/20])

In Examples 7.7 and 7.8, *just* is a focusing premodifier limiting *like* to one point, and it can be glossed to mean *only*. This usage is also identifiable at the level of grammatical prescription, in which *just like* is always followed by *to infinitive*.

Finally, in a very small part of the sample (3.8%) of *just like* occurrences, *like* is associated with its *like*₃ pragmatic function, and it is used as both a filler (*like*_{3.1}, as shown in Example 7.9) and as a discourse marker (*like*_{3.2}, as shown in Example 7.10).

(7.9) What if they **just like** ignore me while I'm
Saying it? (ACAD[Adolescence])

(7.10) And he was **just like**, "Tiffany, look where that
city is." (SPOK[ABC:20/20])

In the remaining sections, I will look at various lexicalized patterns and phrases associated with *just like* to determine whether or not they can be neatly placed within D&L's grammatical framework. I will also refer to the more modern corpus-based *CCLD*, which includes a much more thorough treatment of patterns and phrases than the more traditional M-W dictionary referred to in Section 7.2.

7.4: *Just Like* Confirming Prescribed Rules

7.4.1: Intensifying *Just Like*

7.4.1.1: Just Like Any/Every/All

The most frequent intensifying *just like* construction identified in the corpus data involves the pattern *just like any/every/all*, in which *just like* is followed by one of three distinct determiners, or by one of their related pronoun forms. The main determiners are *any*, *every*, and *all*, and they are often followed by a noun phrase. This construction, although seemingly compositional and transparent, nevertheless takes on a slightly lexicalized or specialized meaning. It is used to show that the person or thing being described is (or behaves) the same as some other group of people or things—implying that it is not unique or special, or that it is average. The construction can be glossed with *the same as*.

A sample of 500 concordance lines for *just like*, taken from the 612-million-word COCA, included 40 occurrences of *just like* followed by the determiners *any*, *every*, and *all*, accounting for eight percent of the sample. A more in-depth search of this construction, using the entire COCA, turned up a total of 19 variations, as shown in Table 7.5.

Table 7.5: Frequency of *just like any*, *just like every*, and *just like all* phrases in the COCA

<i>Just Like</i> Phrase with <i>Every</i>	Freq.	<i>Just Like</i> Phrase with <i>Any</i>	Freq.	<i>Just Like</i> Phrase with <i>All</i>	Freq.
<i>just like everybody</i>	269	<i>just like any other</i>	400	<i>just like all N</i>	73
<i>just like everyone</i>	243	<i>just like any N</i>	166	<i>just like all the other N</i>	51
<i>just like every other N</i>	115	<i>just like anybody</i>	77	<i>just like all of N</i>	36
<i>just like every N</i>	73	<i>just like anyone</i>	73	<i>just like all the others</i>	23
<i>just like everything</i>	71	<i>just like anything</i>	43	<i>just like all the rest</i>	17
<i>just like every-where else</i>	8	<i>just like anywhere</i>	7	<i>just like those N</i>	16
				<i>just like all other N</i>	11
	779		766		227

As can be seen in Table 7.5, the *just like every*, *just like any*, and *just like all* constructions share very similar patterning, with *just like ever* and *just like any* occurring at roughly the same frequency in the COCA (779 and 766 occurrences respectively). The *just like all* construction is less frequent, with 227 total occurrences.

The majority of the *just like every* constructions found in the data consist of the pronouns *everybody* and *everyone*, which account for over 65 percent of all the *just like every* occurrences in the COCA. Examples 7.11 and 7.12 illustrate.

(7.11) If we do something, we get the punishment **just like everybody else** does. (NEWS[WashPost])

(7.12) Universities have limited resources **just like everyone else** these days, and their number one concern remains their students. (MAG[Inc.])

The most frequently occurring of these *just like DETERMINER* patterns with *other*, *just like any other*, occurs 400 times in the COCA. In the vast majority of these occurrences (371), *any other* is followed by a noun, with the four most frequently used nouns being *kid*, *business*, *day*, and *person*. There are 12 occurrences of *kid*, and nine each for *business*, *day*, and *person*. See the following examples:

(7.13) But now he says that Sean is **just like any other kid** on the block. (SPOK[NBC_Dateline])

(7.14) And **just like any other business** they have Operating budgets and equipment requirements. (MAG[Forbes])

(7.15) **Just like any other day**, Dr. Jessica Sevilla Pedraza went to work at the hospital that morning, came home for a quick lunch and then left again. (News[MinneapolisStarTribune])

(7.16) She can do anything she wants when she grows up, **Just like any other person**. (SPOK[Ind_Geraldo])

Also worth noting is a common feature shared between the *just like every* and *just like any* constructions that are used with pronouns rather than determiners: the inclusion of the adverb *else* in the vast majority of all occurrences. This can be seen in Examples 7.11 and 7.12 above for *every*, and in Examples 7.17–7.18 below for *any*. In the case of *just like everywhere*, all eight occurrences in the COCA are followed by *else* (see Example 7.19). Although they are much less common, there are nevertheless some instances where *else* is not used, as can be seen in Example 7.20.

- (7.17) Once you get citizenship, you are **just like anybody else**. (SPOK[CBSFaceTheNation])
- (7.18) And they are made **just like anything else**, through hard work. (ACAD[PlasticSurgery])
- (7.19) Maybe it's because in the kitchen, **just like everywhere else**, we barely have to move anymore! (News[Chicago])
- (7.20) "For 20 years I fought for the rights and dignity of ordinary people, **just like everyone** in this room," Edwards told a standing-room-only crowd. (MAG[WashMonth])

The most frequent *just like all* pattern, *just like all NOUN*, occurs 73 times in the COCA. Again, this pattern functions in the same way as the *just like any* and *just like every* patterns, and it is used to show that someone or something is typical of a group and not unique. Variations of this construction include *just like all the other NOUN* and *just like all the rest* (which, in most cases, consists of *the rest of us*). Examples 7.21–7.23 illustrate.

- (7.21) Because Martin is **just like all the guys you date**. (FIC[Bk:UnderHerSkin])
- (7.22) What do you know, it was **just like all the other courtrooms** I had been in. (SPOK[ABC_Nightline])
- (7.23) He looks **just like all the rest of us**, in his forties, dressed casual-Friday, short hair and the look of a detist trying to relax. (Fic[Esquire])

Now, all the variations of the *just like every/any/all* construction outlined above appear to have a lexicalized component, as they are used to show that a person or thing is average or not unique. However, they can still be accounted for with reference to D&L's grammar. Each variation of this construction can be found in the COCA without the use of *just*. With the lexicalized meaning of *average* or *not unique* intact, such versions show that *just* is being used to intensify the construction. See Examples 7.24–7.26.

- (7.24) I'm going to leave at a specific time **like everybody else** in my office does.
(SPOK[ABC_GMA])
- (7.25) Clinics are for profit, and **like any other business**, they must get and keep clients.
(SPOK[CBS_Sixty])
- (7.26) But I anticipate you'll be like all the other candidates and say it was where you first dreamed of flying into outer space.
(FIC[Analog])

7.4.1.2: *Just Like Always*

The prepositional phrase *just like always* occurs in the COCA 55 times, and while occurring less frequently, it appears to work in a way that is similar to the *just like every/any/all* construction outlined above. That is, it is used to express that something happens or is done in a usual way and is not unique. Grammatically, the main difference between the two patterns is with the part-of-speech used to fill the slot following *just like*. Where the previous pattern takes a determiner or related pronoun, this pattern takes the adverb *always*.

- (7.27) Mike was there, **just like always**, clasping her fingers in his own, pressing his thorny thumb into the hollow of her palm.
(FIC[Bk:SimpleGift])

- (7.28) They buy power from the local utility, **just like always**. (MAG[MotherJones])
- (7.29) Her wish, **just like always**, was his command. (FIC[Bk:MistressNoMore])

This phrase can be glossed with *as usual* and keeps the same meaning. *Just like* can also be glossed with *same* in this phrase and there are 145 instances of *same as always* in the data.

As with the *just like every/any/all* construction, here *just* acts as an intensifier while the shorter phrase *like always* was found to occur 270 times in the data.

- (7.30) La Marque is very balanced on offense, and **like always**, they're aggressive on defense, and their team speed is very good. (NEWS[Houston])
- (7.31) She left for school **like always**, took her lunch pail, and caught the bus at the end of our road about six fifteen. (FIC[Bk:StoneColdDeadEllie])
- (7.32) I figured I'd do it backasshalfwards, you know, **like always**. (MAG[Esquire])

One noticeable difference with this phrase, however, is in its mode of usage: nearly 80 percent of its occurrences (43 out of 55) are found in fiction. The remaining instances were found in newspapers (6) and magazines (4), with only two occurrences coming from the spoken genre subcorpus. This is markedly different from *just like every/any/all*, which is used most frequently in spoken English.

7.4.1.3: *People Just Like You*

People just like you occurs 31 times in the COCA, with the majority of its usage in the spoken genre subcorpus. Here again, *just* is used to convey sameness and it stresses that someone is sharing the same experiences or has the same qualities as a particular group of people.

- (7.33) And we've met **people just like you** all across

- the country, people that never took the yard signs down. (SPOK[CNN_King])
- (7.34) Many are **people just like you**, chasing the American dream. (SPOK[NBC_Dateline])
- (7.35) And the scientists in the lab are **people just like you**, and they are concerned just like you. (SPOK[CNN_Talkback])

The phrase *people like you* occurs 769 times in the COCA, showing that *just* acts as an intensifier here too. It should be noted, however, that most instances of *people just like you* in the data came from TV news or radio programs, reflecting the makeup of the corpus. So, this phrase may not be as common in casual spoken discourse.

7.4.1.4: *Just Like You and Me*

Another related phrase is the more inclusive *just like you and me*, which includes the speaker in the group being compared to. It occurs 41 times in the data and has a wider range than *people just like you*, with 13 instances in the spoken section, 12 in the fiction section, nine instances in newspapers, five in magazines, and two in academic writing. See examples 7.36–7.38.

- (7.36) It is a big day for small investors because starting today average folks **just like you and me** will be able to buy and sell stocks until 8:00 P.M. Eastern Time. (SPOK[Fox_Cavuto])
- (7.37) A lot of those recipes come from people **just like you and me**—more like you, maybe than me. (SPOK[CBS_Morning])
- (7.38) He looked human, **just like you and me**, only better-looking. (FIC[FantasySciFi])

The un-intensified *like you and me* was found 266 times in the data. See Examples 7.39 and 7.40.

- (7.39) That was just something that was invented by people **like you and me** and by the print media, just as a distraction. (NEWS[Chicago])
- (7.40) Why is it that men **like you and me** can take

something good, something well-meaning and virtuous, and squeeze it dry? (MAG[MensHealth])

7.4.1.5: *Just Like PRON SAY*

Variations of the construction *just like PRON SAY* occur 182 times in the data, with 70 instances of the most frequent form, *just like you said*, as shown in Table 7.6.

Table 7.6: Frequency of *just like PRON SAY* phrases in the COCA

Phrase	Frequency
<i>just like you said</i>	70
<i>just like I said</i>	26
<i>just like he said</i>	23
<i>just like they said</i>	13
<i>just like she said</i>	12
<i>just like they say</i>	9
<i>just like it says</i>	7
<i>just like you say</i>	6
<i>just like we said</i>	5
<i>just like I say</i>	3
<i>just like he says</i>	2
<i>just like it said</i>	2
<i>just like she says</i>	1
<i>just like she say</i>	1
<i>just like ye said</i>	1
<i>just like everybody say</i>	1
	182

It is used primarily in spoken English and fiction to mean “I agree with you” or “what you said is true.” See Examples 7.41–7.43.

- (7.41) The ground-up chalk worked **just like you said** it would. (FIC[Analog])
- (7.42) And **just like you** said, Dr. Drew, this is a story about addiction, possibly, maybe there's some mental issues, but she wasn't in her right mind at all. (SPOK[Dr.Drew])
- (7.43) North, away from the lake, **just like you said**. (FIC[Analog])

This construction is an intensification of the *like PRON said*, which occurs 4,602 times in the corpus. See examples 7.44 and 7.45.

- (7.44) **Like you said** it is their fourth album.

(SPOK[NBC:TodayShow])

- (7.45) **Like you said** Jack, I'm just a New York lawyer now. (FIC[Mov:Nixon])

7.4.1.6: *Just Like Old Times*

The expression *just like old times* occurs 63 times in the COCA, with the majority of these occurrences (41) being found in the Fiction section. The remaining occurrences are spread amongst the newspaper, spoken, and magazine sections, with ten, seven, and five occurrences respectively. See Examples 7.46–7.48.

- (7.46) Though they talked easily, **just like old times**, she refused to open up. (FIC[Bk:RestWicked])
 (7.47) And then, **just like old times**, they pulled out the camera. (SPOK[CBS:48Hours])
 (7.48) The Cincinnati Reds are churning out victories, **just like old times**. (NEWS[USAToday])

The similar expressions *just like the old days* and *just like the old times* occur 24 times and just three times, respectively. See Examples 7.49 and 7.50.

- (7.49) **Just like the old days** when we had the booze out here. (SPOK[CBS_Early])
 (7.50) There he was today, glasses on his head after his client was arraigned, feisty, rumped, **just like the old times**. (SPOK[ABC_PrimeTime])

This construction and its variants find use, often nostalgically, to describe something in the present as very similar to something in the past. The unintensified version *like old times* occurs 121 times in the data. It is shown in Examples 7.44 and 7.45.

- (7.51) It was **like old times** with the crowd roaring during the national anthem from the first note and spirits soaring. (NEWS[Chicago])
 (7.52) What she'd really like to do is go sit in a dark bar **like old times**, before he quit. (FIC[SouthernRev])

7.4.1.7: *Just Like NOUN (to VERB)*

The *just like NOUN (to VERB)* construction takes on a more specialized, or lexicalized, meaning than just similarity or sameness—that of typicality—and in all the examples listed below, *just like* can be glossed with *typical of*. It always refers to people, using the form of a personal subject pronoun or a proper noun, and it is used to describe some action or behaviour as typical of the person or people being talked about. This construction occurs a total of 65 times in the COCA. There are 34 instances of *just like PRONOUN (to VERB)* and 31 instances of *just like PROPER NOUN (to VERB)*. See Examples 7.53–7.56.

- (7.53) It was **just like him** to downplay his success.
(FIC[Cosmopolitay])
- (7.54) It would **be just like them** to lie to him, tell him, he was dead, when actually he was just . . . just . . . (FIC[FantasySciFi])
- (7.55) It would be **just like Mike** to go back there to make sure that thing was burning.
(SPOK[NBC_Dateline])
- (7.56) It would be **just like Susan** to suggest they call with the news. (FIC[Bk:UnderBeetles])

While this construction is almost always followed by *to infinitive*, there are few instances in the data with no *to infinitive*. Such an instance is shown in Example 7.57.

- (7.57) And isn't it **just like him**? (ACAD[AmerScholar])

Also, one common variation of this construction involves a negative question form which is used rhetorically, as seen in Example 7.58.

- (7.58) Ain't it **just like you** to believe every word you hear said? (FIC[FantasySciFi])

The *just like NOUN (to VERB)* construction is somewhat different from the other constructions discussed above that serve to illustrate the

grammatical prescriptions for intensifying premodifiers outlined by D&L. In the above constructions, the intensified *just like* form is less frequent than the standard form. However, in the case of the *BE just like NOUN (to VERB)* construction, the intensified version is more than twice as common as the non-intensified version. There are only 26 occurrences of this construction without *just* (16 of them with a pronoun, and 10 with a proper noun), compared to 65 occurrences of the intensified version. This seems to suggest that the intensified version may currently be developing into a more fully lexicalized construction to express typicality.

7.4.1.8: Sense Verbs and *Just Like*

One notable finding with *just* as an intensifier is that when paired with the polysemous perception verbs (PV), there seems to be an effect on the frequency distributions of the various senses in which the *PV like* construction is used. This effect shows mostly where the PV is used as a true sense verb. The most notable change occurs with *LOOK like*: In Chapter 6, I found that *LOOK like* is used for visual sense description 67 percent of the time and for situational evaluation 33 percent of the time (Section 6.4). When *LOOK like* is premodified with *just*, however, this distribution of usage shifts dramatically, with nearly all occurrences being used for visual sense description. There are 1,181 occurrences of *LOOK just like* in the COCA, and in a 500-concordance-line sample, all but two lines were identified as being used for visual sense description. See Examples 7.59–7.61.

(7.59) You **look just like** my library teacher.
(SPOK[NPR_ATCW])

(7.60) There's a lot of fake stuff out there that **looks just like** leather, and it doesn't cause animals to suffer and die. (News[Atlanta])

(7.61) You **look just like** you know what you're doing.
(FIC[Bk:TrueBetrayals])

The frequency distribution of the four senses of *FEEL like*, as reported in Chapter 6, showed that *FEEL like* is used most of all for situational evaluation, at 50 percent of the occurrences. I found that it describes feelings 33 percent of the time, describes moods/wants (*like*_{2.3}) ten percent of the time, and describes tactile sense only six percent of the time. Now, there are only 118 occurrences of *FEEL just like* in the COCA, but when I analyzed the data, 60 occurrences (50.8%) showed use for describing feelings. Another 37 occurrences (31.4%) were used for describing tactile sense, and 21 (17.8%) were used for situational evaluation. I found no instances of *FEEL just like*, used to mean *in the mood for* (*like*_{2.3}), in the data. See Examples 7.62–7.64.

(7.62) You know, **I feel just like** a piece of meat here.
(SPOK[Ind_Limbaugh])

(7.63) “The result is a surface that **feels just like** sandpaper,” she describes. (MAG[AmerArtist])

(7.64) And after a certain point, I think you kind of forget as you’re watching it, and you just get so into the story, and it **feels just like** a natural love story. (SPOK[PBS_Newshour])

Similarly, *SOUND like* tends to skew more toward its true sense-verb usage when premodified with intensifying *just*. As noted in Section 6.6, *SOUND like* was used for situational evaluation in 53 percent of the data sample, and for aural sense description in 47 percent of the sample. Of the 254 occurrences of *SOUND just like* in the COCA, 222 (87.4%) were used for aural sense description, and the remaining 32 occurrences (12.6%) were used for situational evaluation. This finding needs to be interpreted with caution, however, as there are many instances of *SOUND just like* that can be construed as either situational evaluation or sense description. As mentioned in Chapter 6, the boundary between the various senses of the *PV like* construction can be rather fuzzy at times, and this turned out to be

particularly true when I was examining the data in this area. Example 7.65 could be seen as either aural sense, as in the speaker and their father have similar sounding voices and speaking content. Or it could be seen as situational evaluation, where the speaker's behaviour is being compared to their father's, based on what they are saying (or have said). In example 7.66, on the other hand, it seems quite clear that the speaker is talking about how their singing literally sounds as if they could compare themselves to a particular, famous singer.

(7.65) I **sounded just like** my father.

(FIC[Bk:Left-handWay])

(7.66) I can **sound just like** Celine Dion if I want to,
says Thom. (MAG[Time])

While this is an interesting interaction involving *just* and its effect in the polysemous *PV like* construction, it is not exactly clear why *just like* tends to be used to intensify sense description more often than situational evaluation. However, it is clear here that *just* shows the behaviour of an intensifying premodifier, and that it follows the traditional grammatical framework of intensification as outlined by D&L.

7.4.2: Focusing *Just Like*

As previously mentioned in Section 7.3.2, less than six percent (5.8%) of all *just like* usage in the COCA is verbal *like*₂ usage. The data from the sample suggest that when verbal *like* is premodified with *just*, it always functions as a focusing premodifier. In these cases, *just* is used to mean *only*, and this meaning is listed as sense 3a in the M-W; it is used pragmatically, as a polite or hedged form. The corpus data also suggest that this form of *just like* is used to mean *only*.

But functioning as a polite form does not appear to yield nearly as many lexicalized patterns or phrases in the same way as the intensifying

just like patterns and phrases outlined in Section 7.3 above. For example, the most frequent recurring pattern with focusing *just like*, *just like to VERB*, is fully compositional and does not take a lexicalized meaning in the same way that *just like NOUN to VERB* is used to express typicality. However, there are several recurring verbs that are used in the *just like to VERB* pattern, as shown in Table 7.7.

Table 7.7: Most frequent verbs found in the focusing *just like to VERB* pattern in the COCA. Only verbs with frequencies of 10 or above are included.

VERB	Frequency
say	184
know	72
ask	59
see	55
make	39
get	37
go	25
thank	19
point	16
talk	14
play	13
look	12
hear	12
add	12
take	12
remind	12
tell	10
think	10

As can be seen in Table 7.7, the most frequent *just like to VERB* pattern found in the COCA is *just like to say*, with 184 occurrences. Now, this appears to be a fully compositional construction, and it follows D&L's framework, with *like* classified as focusing premodifier, which can be glossed to mean *only*.

However, there is a somewhat specialized, pragmatic, function that can be seen in Examples 7.67 and 7.68, where the *would just like to say* construction is used to show politeness.

(7.67) On behalf of us all, I'd **just like to say** a

- great big thank you to our secretary, Freda Kelly. (SPOK[CBS:NewsSundayMorning])
- (7.68) I would **just like to say** good luck to everybody today and I hope you guys run a – run well. (SPOK[CBS_SunMorn])

Examples 7.69 and 7.70, on the other hand, do not carry the same sense of politeness as in Examples 7.67 and 7.68. In 7.69, the subject is *you* rather than *I*, and the modal *would* is not used. In 7.70, there is no focusing *just*—and, here too, the function of politeness is missing. The evidence suggests that *I would just like to say* has taken on a semi-lexicalized function.

- (7.69) You **just like to say** it does! (NEWS[SanFranChron])
- (7.70) In closing, I'd **like to say** directly to the president, we honor our veterans, even in the rain. (SPOK[Fox_Five])

The other recurring verbs in the *just like to VERB* construction appear to function as outlined in D&L's grammatical classification, with *just* functioning as a focusing premodifier to mean *only*. This can be seen in Examples 7.71–7.73 below.

- (7.71) Take all the time you want. I would **just like to know** your plans so I can plan accordingly. (FIC[NewYorker])
- (7.72) We'd **just like to ask** you a few questions, if we could. (FIC[Mov:AnalyseThat])
- (7.73) I **just like to see** him. (SPOK[CNN:AndersonCooper])

7.5: *Just Like* Contradicting Prescribed Rules

Section 7.4 above demonstrates several frequently occurring constructions based on *just like* that confirm the grammatical prescription for how to use both intensifying and focusing premodifiers as outlined by D&L. In the data,

however, I identified an additional three *just like* constructions that do not fit within this grammatical framework.

7.5.1: *And Just Like That*

The phrase *and just like that* occurs 198 times in the COCA and is used to show that something occurs *quickly*, *suddenly*, or *easily*, as shown in Examples 7.74–7.76 below.

(7.74) **And just like that** he was done. (MAG[Essence])

(7.75) That was . . . **and just like that** she fainted.
(FIC[Bk:DeadlyRuseMacMc])

(7.76) **And just like that** the machine ran with a new
steady hum, just like it had when it was new.
(FIC[Bk:DamnedIfIDo])

Looking at these examples, we can see that *just* does not function as an intensifying premodifier. This is because *just like* cannot be glossed with *same as*. And, as *just like* cannot be glossed with *only*, the phrase does not function as a focusing premodifier.

While the majority of occurrences of this construction are formed with *and*, an alternate form of this construction was found using the pattern *VERB just like that*, as shown in examples 7.77–7.79. This form is much less frequent. The main verbs used in the pattern are *HAPPEN*, *CHANGE*, and so forth and these verbs all seem to belong to a similar semantic set.

(7.77) You know, 12 years from my life, it's **gone just like that**, because of this regime.
(SPOK[CBS_48Hours])

(7.78) The time **changes just like that**. (FIC[Esquire])

(7.79) Mommy got sick and it **happened just like that**
and there was nothing anybody would do.
(FIC[Mov:AmericanPresident])

Further exemplifying the lexicalized nature of *and just like that*, instances of *just like that* in the COCA—without *and*—do not take the meaning of *quickly* or *suddenly*. Instead, *just* performs its standard grammatical functions of

intensification (see Examples 7.80 and 7.81) and, less frequently, focusing (with verbal *like* as seen in Example 7.82). In the case of Example 7.73, the meaning of *just like* was not immediately clear, and could be construed either as meaning *suddenly* or as an intensified form. Analysis of the co-text, however, revealed that this is taken from a court transcript in which a lawyer is questioning a witness and *just like that* is acting as an intensified deictic form.

(7.80) You hit him **just like that**? (SPOK[CBS_48Hours])

(7.81) He's **just like** that grasshopper
(FIC[SouthernRev])

(7.82) Women **just like that** tough image . . .to go for
the biggest, the strongest, the toughest, and
that's what the bad boys portray. (MAG[Ebony])

Furthermore, the shorter form *like that*, which was found to occur very frequently in the data (60,413 occurrences), is almost always functioning as a deictic, and is not used to mean *quickly* or *suddenly*. See Examples 7.83 and 7.84.

(7.83) Then we expect them not to act **like that**?
(NEWS[USAToday])

(7.84) Did Elsie **like that**? (FIC[NewEnglandRev])

The evidence outlined here shows that the phrase *and just like that*, as well as the related pattern *VERB just like that*, when used with a small set of related verbs, will show a specialized meaning of *quickly* or *suddenly*. It is a lexicalized construction that has not been explained using the traditional dictionary-and-grammar model of language.

7.5.2: Pragmatic Use of *Like*₃

In spoken English, *like* is used pragmatically as both a filler (*like*_{3.1}) and a discourse marker (*like*_{3.2}). Both of these functions of *like* were found to co-occur with *just*. Examples 7.85 and 7.86 show *just like* being used as a filler,

and Examples 7.87 and 7.88 show *just like* being used as a discourse marker.

- (7.85) And I mean it w—it was **just like**—I didn't know what to do, and I just, you know, started shaking and my heart is pounding, and—and I said, Well, where—wh—where is he?
(SPOK[Ind_Geraldo])
- (7.86) What if they **just like** ignore me while I'm saying it? (ACAD[Adolescence])
- (7.87) And then afterwards, I was **just like**, I can't—I can not cry. (SPOK[CBS_Morning])
- (7.88) And he was **just like**, "Tiffany, look where that city is." (SPOK[ABC:20/20])

In Examples 7.85 to 7.88 above, *just* functions as neither an intensifying premodifier nor a focusing premodifier. This pragmatic usage of *just like* does not fit with D&L's grammatical framework, and there is no mention in the grammar of such pragmatic usage for *just like*.

Additionally, the extended phrase *just like, you know*, is used as a filler (see Examples 7.89 and 7.90) and, in Example 7.91, it appears to be used as a discourse marker. It occurs 92 times in the COCA, with 69 of these instances in the spoken section of the corpus. It is clear, from this usage also, that *just* does not act as an intensifying premodifier or as a focusing premodifier.

- (7.89) It's not **just like, you know**, all-out debauchery.
(SPOK[NBC_Dateline])
- (7.90) He was **just like, you know**, I mean, God, man, that's amazing. (SPOK[NBC_Today])
- (7.91) It was **just like, you know**, Be careful.
(SPOK[BC_Dateline])

7.6: Representation of *Just* and *Like* in the CCLD

I would like to return to the point of view of a language learner trying to find the meaning of *just like*. Where I left off in Section 7.2, I provided an example of *just like*, taken from the COCA. I will place it here for convenience, as I wish to discuss what information can be found in a learner's dictionary.

(7.1) His front and back legs moved in unison, **just like** a bear's. (FIC[SouthernRev])

In Section 7.2 I outlined the lexical representation of both *just* and *like* as shown in the M-W. While this dictionary did provide a list of phrases associated with both *just* and *like*, however limited, the main problem was that only one phrase could be found consisting of *just like*: *I'd just like to see him try*. This was not the meaning that our language learner was looking for. Furthermore, a search of the COCA for that pattern, *just like to see PRONOUN try*, showed that it does not occur at all in the corpus. So, from a lexical syllabus perspective, it would not appear to be a very useful phrase for a learner. And based on this finding, we can say that the mainly traditional dictionary was not very helpful in determining the meaning of *just like* that we found in Example 7.1. How then, might a more contemporary, corpus-based dictionary perform? One such as the CCLD can serve as a good example.

The first thing our language learner would notice when looking up the target words in the CCLD is the vast amount of information given for each word. There are two main entries for *just*, with 23 senses listed for *just* as an adverb, and one for *just* as an adjective. There are also main entries for *like*, with 19 senses listed for preposition and conjunction uses, and another 12 senses listed for verb and noun uses. While this is a lot of information to look through, the CCLD includes the pattern grammar associated with each

sense in the margin, and this can be very helpful when searching for the right sense of a word.

An examination of all the information found in the entries for both *just* and *like* reveals two *just like* constructions: Sense 18 of the adverbial uses of *just* provides the following account:

You use **just** in expressions such as **just like**, **just as . . . as**, and **just the same** when you are emphasizing the similarity between two things or two people. *Behind the façade they are just like the rest of us . . . He worked just as hard as anyone.* (600)

And Sense 6 of *like*, used as a preposition and conjunction, states:

You can use *like* in expressions such as **that's just like her** and **it wasn't like him** to indicate that the person's behaviour is or is not typical of their character. *Why does he want to do a mad thing like that? It's not like him.* (637)

The inclusion of these *just like* expressions in the *CCLD* are very helpful, and Sense 18 of *just*, shown above, is exactly the meaning of *just like* that our language learner is looking for. This would help the learner to understand the sentence shown in Example 7.1. Furthermore, Sense 6 of *like* corresponds to the *just like NOUN (to VERB)* construction outlined in Section 7.4.1.7. It seems, then, that a current corpus-based dictionary can be more helpful in the search to understand multi-word constructions. That is, it can be more helpful than either a more traditional dictionary or a traditional grammar. However, the *CCLD* only provides definitions for a small fraction of the *just like* constructions found in the data. And while the *CCLD* entries would likely help the learner understand the more compositional constructions discussed above, it is problematic that the most lexicalized constructions are not found in the *CCLD*. So, those lexicalized constructions covered in neither the traditional dictionary nor the traditional grammar—(and) *just like that* as well as pragmatic uses of *just like*—would remain a mystery.

7.7: Conclusion

In this chapter, through an examination of the *just like* construction, I have demonstrated the problem with traditional lexical and grammatical classification at the single-word level.

I have examined the most frequent patterns and phrases consisting of *just like*, and I have compared these findings to the lexical and grammatical classifications of *just* and *like* in a traditional dictionary and grammar. I have shown how the corresponding dictionary-and-grammar model of language can, at times, prove insufficient or even misleading when dealing with multi-word lexical constructions. Furthermore, a more current, corpus-based dictionary was found to be more helpful than the traditional dictionary in identifying the most common use of *just like*. And yet, other lexicalized uses of this construction were not found in the corpus-based dictionary. This shows how complex the phraseology of a highly-frequent word such as *like* can be.

8

Results

Like with Comparative and Superlative Premodifiers

8.1: *Like*, Prescription, and Description

In this chapter, I will examine the corpus data to determine how prepositional *like* functions when used with comparative and superlative premodifiers. By comparing these corpus findings with a more traditional, prescriptive grammatical explanation of comparative and superlative forms, I will further demonstrate the problems that can arise, for teachers and learners, when explanations rely on such traditional lexical and grammatical classification. Continuing with my case study of *like*, I will show that the grammatical account of comparative and superlative forms is inadequate because it does not fully address all the uses of comparative and superlative forms used with *like*. That is, the contrast between traditional prescriptions and corpus-based descriptions can reveal that the former are sometimes misleading due to their oversimplification, and this is especially true when phraseology is taken into account.

This chapter is organized in four main sections, 8.2–8.5. In Section 8.2, I will review the traditional grammatical prescription concerning comparative and superlative degrees, as used in English. I will also provide frequency data for comparative and superlative lexical items that use *like* in

the Corpus of Contemporary American English (COCA). In the remaining three sections after that, I will provide a detailed analysis of each comparative and superlative lexical item used with *like*, and I will compare the corpus data with the grammatical description outlined in 8.2. First, in 8.3 I will identify the constructions found in the data that are included in Downing and Locke's (2006; D&L) grammatical analysis of comparative and superlative forms. Next, in 8.4, I will identify and analyze the constructions found in the data that are grammatically based but not included in D&L's grammar. And, finally, in 8.5 I will examine two fully lexicalized phrases that cannot be accounted for in a traditional grammatical description.

8.2: Comparative and Superlative Premodifiers

Table 8.1, reproduced from D&L (486), shows the authors' grammatical summary of both the inflectional and the analytic grading options used with adjectives to express comparative and superlative degrees.

Table 8.1: Summary of inflectional and analytic comparative and superlative grading options for adjectives as described in a traditional grammar resource. Reprinted from Downing and Locke, *English Grammar: A University Course*, 2006.

Scale of degree	Inflectional	Analytic
1. Comparative superiority 2. Superlative superiority 3. Equality 4. Comparative inferiority 5. Superlative inferiority 6. Sufficiency	easier the easiest	more difficult the most difficult as easy, as difficult less easy, less difficult the least easy, the least difficult easy enough, difficult enough

In their prescription for this issue in English grammar, D&L explain that comparative and superlative degrees are expressed in two ways. Gradable adjectives that consist of one syllable, or of two syllables where the ending is *-y*, can be inflected with either *-er* or *-est*. Meanwhile, adjectives consisting of two or more syllables, and adjectives that are already inflected (e.g.,

lovable, famous), are graded analytically, most commonly with the adverbs *more, most, less, and least*. This description reflects the typical grammatical explanation of comparative and superlative usage, an explanation typically found in pedagogic materials such as ELT coursebooks. In addition to comparative and superlative degrees of inferiority (*less, least*) and superiority (*-er, -est, more, most*), D&L (484–487) include the analytic forms for expressing equality and sufficiency in their summary. These are sometimes not included with explanations of comparative and superlative usage, as found in basic pedagogic grammars (e.g., *World English* [Martin Milner 2012] or *Four Corners* [Richards and Bohlke 2012]).

While the inflectional rules outlined in Table 8.1 apply only to adjectives, the analytic premodifiers can also be used for comparative and superlative grading of nouns, adverbs, and prepositions. However, D&L note that comparative and superlative grading of prepositions is more restricted and only works with certain prepositions. It is noteworthy that two of the three examples of gradable prepositions provided by D&L consist of *PV like* constructions (discussed in Chapter 7): *seemed more like*, and *not sound in the least like*. This gives the reader the impression that prepositional *like* is frequently used in this way and although they do not provide any frequency data, according to the COCA it is an accurate impression. Additionally, D&L provide an attested list of grading modifiers used with prepositions, including, *more, less, far more, much less, the most, the least, and in the least* (2006, 538).

8.2.1: Comparative and Superlative Found with *Like*

Table 8.2 shows the frequency data for the comparative and superlative forms used with prepositional *like*, as found in the COCA.

Table 8.2: Frequency data for all comparative and superlative premodifiers used with prepositional *like* in the Corpus of Contemporary American English (COCA). Four items are not covered by D&L's (2006) description of comparative and superlative forms (+); *like enough* is not used to express sufficiency in the COCA (*).

Comparative	Freq.	Superlative	Freq.	Equality	Freq.	Sufficiency	Freq.
more	12,643	most	204	as much like+	98	enough like+	101
less	799	the most	58			like enough*	89
more and more	296	least	35				
less and less	27	the least	18				
more or less+	72	at least+	24				

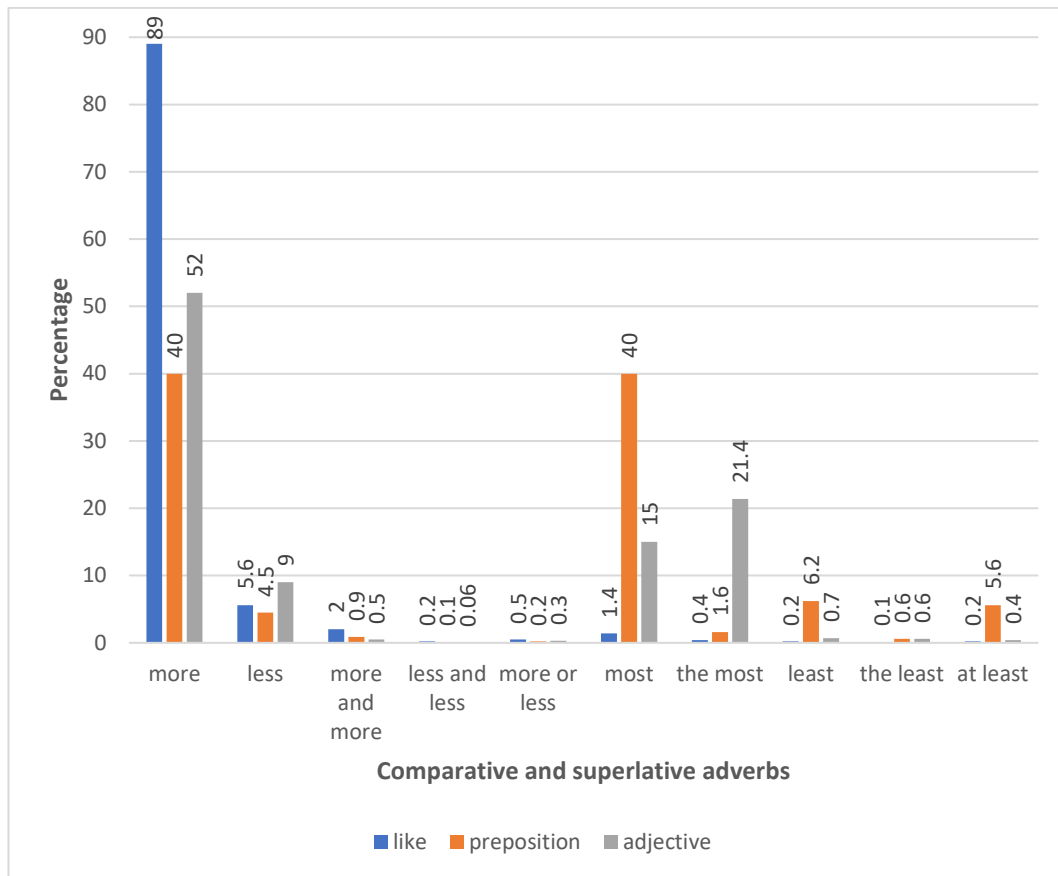
These data show that *like* is used with a wide range of comparative and superlative grading modifiers that includes most of the items in D&L's analysis (Table 8.1; D&L, 486), including all the attested items used with prepositions. Additionally, however, the data present two items that are not mentioned at all in D&L's grammatical analysis. We can also see that the comparative adverbs *more* and *less* are used with *like* more frequently than their superlative counterparts *most* and *least* by a considerably large margin. Most notable here is the dominance of *more*, which accounts for 89 percent of all comparative and superlative usage with *like* in the COCA. Considering this high frequency, it is not surprising to find that *more like* was found to be the most phraseological of the items in Table 8.2 (see 8.4.3.1, 8.5.1, and 8.5.2).

Comparing the corpus findings shown in Table 8.2 with the comparative and superlative grading scale provided by D&L, there are four clear discrepancies. The first two of these involve phrases identified in the corpus data that were not covered in the grammar. The first phrase, *more or less like*, occurs 75 times in the COCA, and the second, *at least like*, occurs 24 times. These two phrases will be discussed in more detail in Sections 8.5.2 and 8.4.3.1 respectively. The remaining two discrepancies are grammatical in nature, and they are the result of D&L's grammar explanation being

incomplete, especially in its prescription for how comparative and superlative premodifiers are to be used with prepositions (2006, 486).

The first grammatical discrepancy concerns the expression of equality, in which the corpus data show that while prepositions do not work with *as* to show equality, the adverb *much* can be paired with *as* to show equality using a preposition. There are 98 occurrences of *as much like* in the COCA, and this construction is not mentioned in D&L's summary. In addition, the lexicalized phrase *about as much like* was identified in the data, and it takes on a specialized meaning. See Sections 8.4.1 and 8.4.3.3 respectively for discussions of these two constructions. Finally, the fourth discrepancy involves showing sufficiency (see Section 8.4.2). While D&L's grammar stated that sufficiency is expressed with the *adjective + enough* construction when grading adjectives, the corpus data show this does not work in the same way with prepositions, as *like enough* does not express sufficiency. However, it was found that the reverse, *enough like*, does work to express sufficiency, with 101 occurrences in the COCA. This, again, differed from the account offered by D&L (2006, 486).

Comparing the percentage distribution of comparative and superlative *like* phrases with the normal distributions of comparative and superlative preposition and adjective phrases (see Figure 8.1), we can see that *like* behaves differently from other premodifiers used in these comparative and superlative phrases.

Figure 8.1: Comparison of percentage distributions for comparative and superlative adverbs used with *like*, prepositions, and adjectives

First, the data in Figure 8.1 show that *like* is used with *more* disproportionately more frequently than with other prepositions, while *most* is used with *like* disproportionately less than the normal distribution of prepositions. In addition, *less* is used with *like* slightly more than other prepositions, and we can see that *least* is used very infrequently with *like* compared with the normal distribution of prepositions. Taken together, these data confirm that *more like* is highly frequent (seemingly at the expense of other comparative and superlative *PREMODIFIER like* phrases), and that it behaves differently from other prepositions in comparative and superlative constructions. Figure 8.1 also shows that the repeated comparative constructions, and the phrase *more or less*, while relatively infrequent, are used more often with

like. This seems to suggest that *like* is used more often than other prepositions in comparative and superlative phrases. The notable exception here is the phrase *at least like* (discussed in Section 8.4.3.2), which is used with other prepositions much more frequently than with *like*.

In the following sections, I will examine the corpus data to determine whether or not they support D&L's (2006) grammatical summary of comparative and superlative forms. I will also determine if there are any further data that do not fit within their account of the grammatical constructions. The main data consist of 500-concordance-line samples for *more like*, along with samples for *less like*, and all instances of *most like*, as well as of *least like*. The latter constructions, *most like* and *least like*, occur less than 500 times.

8.3: Constructions Confirming Prescribed Rules

8.3.1: *More Like*

The data in Table 8.2 and Figure 8.1 show that the comparative construction *more like* is highly frequent in the COCA, occurring 12,643 times (22.4 times per million words). *More* is the most frequently used comparative/superlative premodifier with *like*, and *like* is used in the *more + preposition* construction proportionately more often than other prepositions. In the sample of 500 concordance lines, *more like* was predominantly used for simple comparisons, accounting for nearly 91.6 percent of the sample.

In Examples 8.1–8.6 below, it can be seen that the same base pattern, *NOUN PREMODIFIER like NOUN*, is used to form standard compositional comparisons. Thus, it fits neatly with D&L's (2006) grammatical prescription for comparative and superlative grading. These comparisons include *than*, followed by a complement in 21.6 percent of the sample. Examples 8.1–8.3 provide typical examples from the corpus data for this usage.

- (8.1) At first, the young woman's face had seemed **more like** a doll's **than** a human's
(FIC[Bk: PlainKilling])
- (8.2) Even with an evolutionary makeover, the design looks **more like** a concept **than** a production car.
(MAG[PopMech])
- (8.3) Human beings are far **more like** chips **than** like assembly lines (ADAD[CATOJournal])

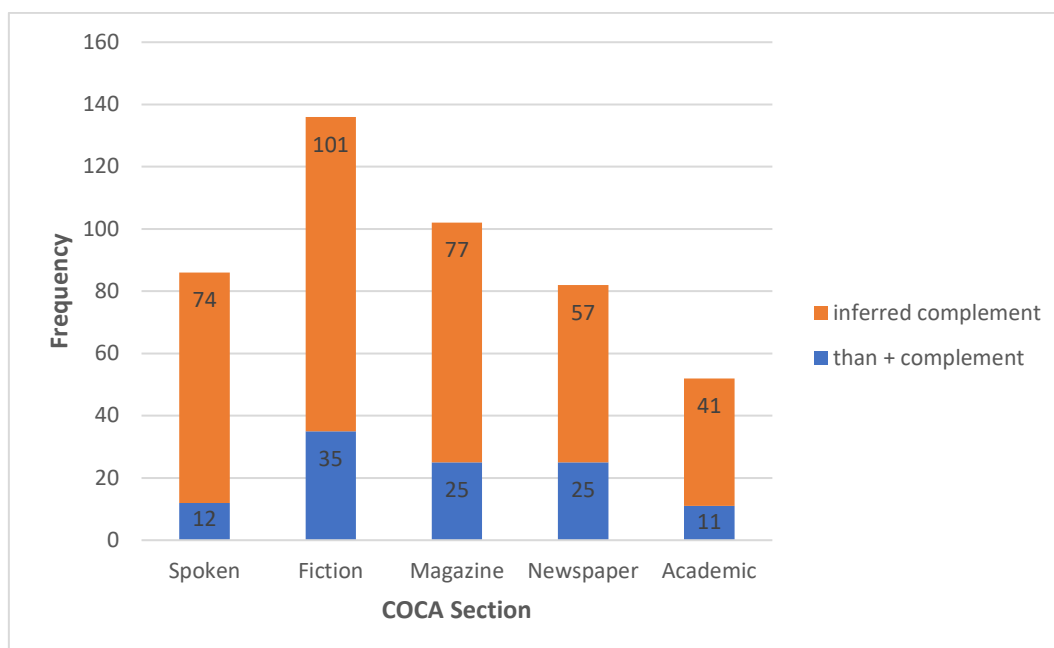
At 70 percent, the majority of *more like* comparisons in the sample however consist of comparative forms used without a complement, as can be seen in Examples 8.4–8.6. In these cases, according to D&L (2006, 487), the other entity in the comparison should be inferable.

- (8.4) Backed by raucous cheers, his concession speech sounded **more like** a victory celebration.
(SPOK[NBC_Today])
- (8.5) "Why can't you be **more like** your sister Anne?" he often shouted at Charles. (MAG[GoodHousekeeping])
- (8.6) They summoned him to a meeting. It was **more like** a trial. (FIC[Read])

Examples 8.4–8.6 are representative of comparative statements in the sample that do not use the *than* COMPLEMENT construction, and in which the missing complement is easily inferable. In Example 8.4 we can see that the complement to the comparative *more like* is *concession speech*. Here, we can see that including a *than* COMPLEMENT construction would be redundant, presenting a concession speech that sounded more like a victory celebration *than a concession speech*. In Example 8.5, it can be inferred that the complement to the comparison is *than yourself* (*Why can't you be more like your sister Anne than yourself?*). Finally, in Example 8.6 the entity being compared is stated in the preceding sentence. The resulting comparison can be understood as follows: *It was more like a trial than a meeting*.

Figure 8.2 shows the frequency data for the comparative usage of *more like* in each section of the COCA.

Figure 8.2: Frequency of comparative *more like* in each section of the COCA, showing separate frequencies for comparisons using *than* COMPLEMENT and for inferred complements (without *than*)



These data show that the distribution of *more like* usage across genres is consistent with the frequency of usage for *like* in general in the COCA. It is most common in fiction by a significant margin, followed by magazines, and then spoken English, its usage in academic English being the least frequent. This consistently high frequency of use in fiction can likely be attributed to writers relying on *like* for its descriptive nature, whether it be for forming similes (as noted in Chapters 5 and 6) or for describing someone or something using comparative or superlative constructions.

8.3.2: *Less Like*

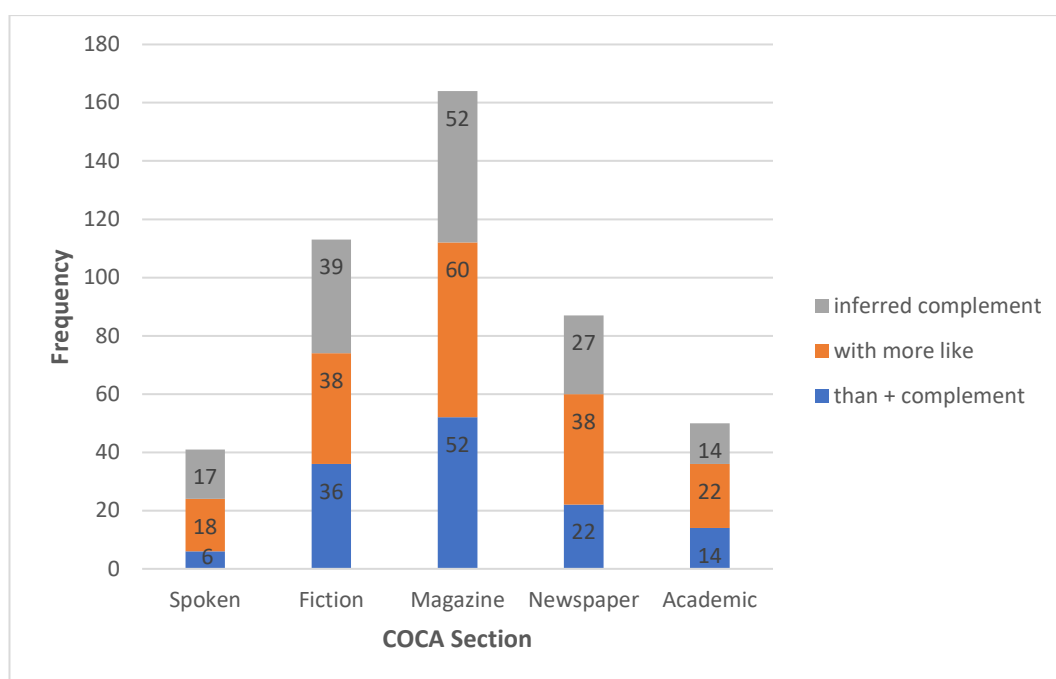
The comparative inferior *less like* occurs 826 times in the COCA. Just over 30 percent of the sample consisted of *less like* being used in simple comparisons without complements (see Example 8.7), and 25.8 percent consisted of comparative statements that included *than* followed by a

complement (see Example 8.8). However, the frequency data for *less like* is misleading. In 35 percent of the sample, *less like* is used in conjunction with *more like* to describe a more dynamic comparison (as in Example 8.9), with the focus being on *more like*.

- (8.7) They make him appear **less like** a college kid,
though he thinks he looks “dorky” wearing them.
(NEWS[Atlanta])
- (8.8) What she had was **less like** a forgetting than an
absence. (FIC[FantasySciFi])
- (8.9) We often feel **less like** a nation and **more like** a
collection of folks venting. (NEWS[CSMonitor])

Usage of *less like* was found to generally conform to D&L’s (2006) grammatical description of comparative usage, using the same grammatical patterning as *more like*. Interestingly, however, the distributions of *less like* among the five COCA sections differ from the normal distribution of *more like*, and from that for *like* in general, as shown in Figure 8.3.

Figure 8.3: Frequency of comparative *less like* in each section of the COCA, showing separate frequencies for comparisons using *than COMPLEMENT* and inferred complements (without *than*), and comparisons paired with *more like*



The data in Figure 8.3 show that the comparative *less like* is used most often in magazines, whereas *more like* is used most often in fiction. The reason for this difference is unclear; however, the most common use of *less like* in this genre involves pairing *less like* with *more like*, as shown in Example 8.9. These comparisons appear more descriptive and dynamic. This observation fits with previous findings concerning magazine English. As shown in Chapter 5, magazines use more verbal predicative similes, as these constructions result in more dynamic comparisons.

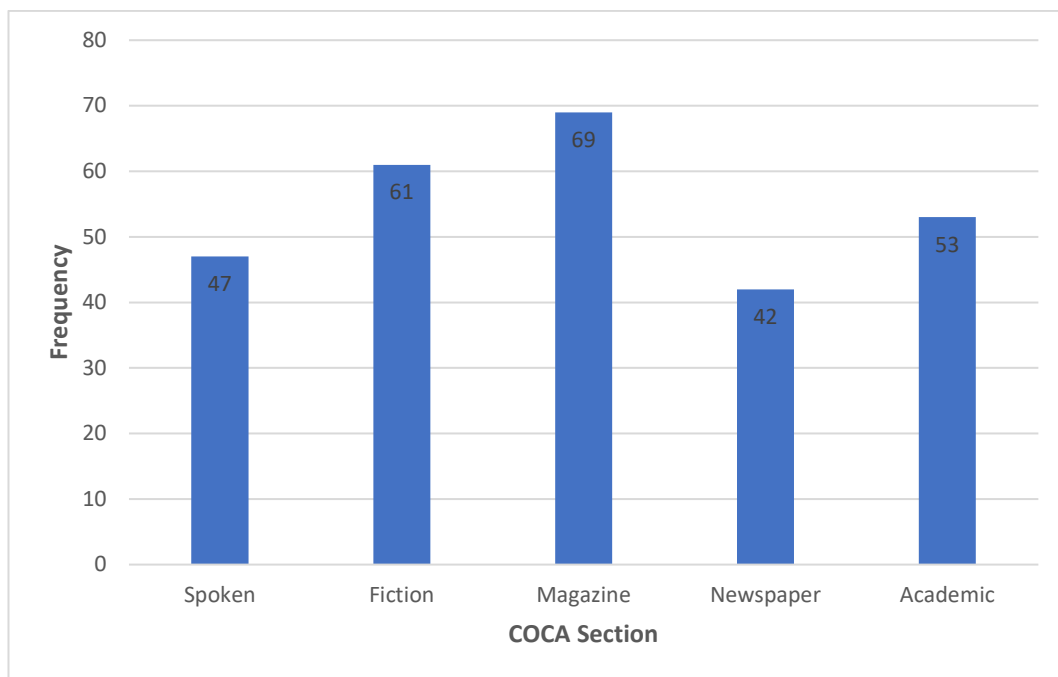
8.3.3: *Most Like*

Over half of the 484 occurrences of *most like* in the COCA (55 percent, or 257 occurrences) involve the prepositional use of *like* to produce superlative statements.¹ These statements are generally compositional, and they follow the same grammatical patterning as the comparative constructions *more like* and *less like*, as shown in Examples 8.10–8.12 below.

- (8.10) Which one of you is **most like** your father?
(SPOK[ABC: 20/20])
- (8.11) Of the great apes-bonobos, gorillas, orangutans,
and chimpanzees-bonobos are the **most like**
humans. (MAG[NatGeog])
- (8.12) In its molding of space through planes of sheet
metal supported by a framed structure, molue
building is **most like** architecture.
(ACAD[AfricanArts])

Figure 8.4 shows the number of occurrences of the comparative construction *most like* for each section of the COCA.

1. Of the sample, 204 occurrences (41 percent) were concerned with the verbal use of *like*. Additionally, there were 11 occurrences of *most like*, seemingly used adjectively as a shortened version of *most likely*, and four occurrences of *like* being used as a discourse filler when following *most*.

Figure 8.4: Frequency of superlative *most like* in each section of the COCA

These data show that usage of the superlative construction *most like* is more evenly spread across all sections of the COCA, with a notable increase in usage in academic English, when compared to the other *preposition + like* frequency distributions. Perhaps this reflects a tendency in academic writing to use superlatives when describing processes such as data collection and the reporting of results, as in Examples 8.13 and 8.14.

- (8.13) Items are in a structured format in which the subject first indicates which of the two types of individuals he/she was **most like** and then whether this likeness was “sort of” or “really” true. (ACAD[Adolescence])
- (8.14) Ask students to listen, with their knowledge of plainchant performance practice, and determine which performance has the most chantlike character (steady tempo, smooth movement, plain singing) and a performance of dissonances that is **most like** their decision from Step 6. (ACAD[MusicEduc])

8.3.4: *Least Like*

There are only 60 instances of prepositional *like* premodified with *least* in the COCA.¹ Of these, there are 23 occurrences of the phrase *at least like* (discussed separately in Section 8.4.3.2), and another 10 occurrences of the construction *in the least*, which was briefly mentioned by D&L (2006, 538). However, when used with *like*, this construction is always preceded by *not*, forming the lexicalized phrase *not in the least like*, and this will be discussed separately in Section 8.3.4.1 below. This leaves only 27 occurrences of *least like* being used compositionally to form a comparative in the COCA, as shown in examples 8.15–8.17.

- (8.15) I think she's very funny, and she's the **least like** me. (SPOK[CBS_SunMorn])
- (8.16) Of all creatures the snake is, of course, the one **least like** us. (FIC[Salmagundi])
- (8.17) Being a hybrid, the Forester is by far the quickest in the group—due to the power output—and feels the **least like** a truck on the road. (MAG[PopScience])

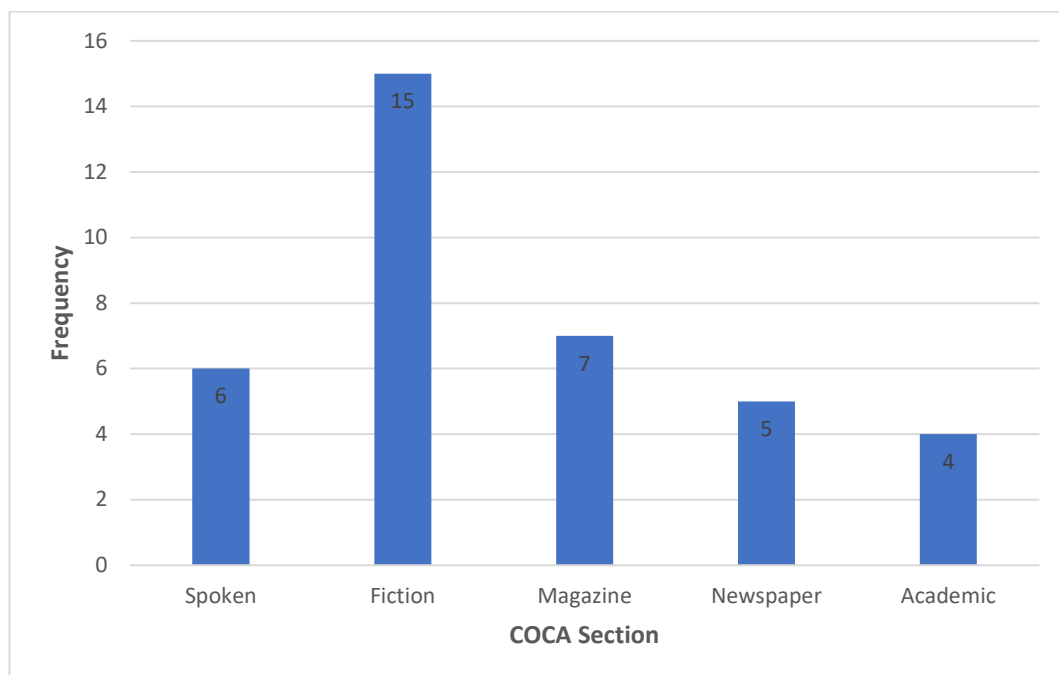
These examples further demonstrate how the usage of comparative and superlative premodifiers with *like*, while it demonstrates the traditional grammatical rules, tends to be used compositionally following the *NOUN PREMODIFIER like NOUN* pattern.

Figure 8.5 shows the frequency of the grammatical superlative *least like*, as found in the five sections of the COCA. The frequency distributions in Figure 8.5 show that the majority of occurrences of *least like* are in fiction. Also, we can see that the pattern of frequency distributions among the

¹ There were 112 total instances of *least like* in the COCA. A large percentage, however, involved the verbal function of *like*, and another two occurrences involved the discourse filler function of *like*.

sections looks very different compared with the pattern of frequencies for *more like*.

Figure 8.5: Frequency of superlative *least like* in each section of the COCA



8.3.4.1: (Not) *In the Least Like*

While D&L include the phrase *in the least* in their list of attested grading modifiers known to work with prepositions, they seem to have missed an important feature of this construction. All ten occurrences of *in the least like*, as found in the COCA, are preceded by a negative, and they form the phrase *not in the least like*. Nine of these ten were found in the fiction section, and one was found in the news section. While this usage is grammatically similar to the comparative and superlative rules outlined by D&L, the phrase *not in the least like* seems to take on the meaning of *not at all like*.

(8.19) On second thought, she **didn't look in the least like** Bessa. (FIC[BkSciFi: QuantumLeap])

- (8.20) Why, it's **not in the least like** a ship.
(FIC[Analog])
- (8.21) But they are **not in the least like** jewelry
enlarged. (NEWS[CSMonitor])

Further analysis of this usage of the *in the least* construction as found in the COCA shows a total of 59 occurrences in which it is modifying a preposition. I found that all but five of these occurrences involve the negative construction, used to mean *not at all*. It seems like a significant, potentially misleading oversight by D&L to not mention the negative element of this construction. Further, it is all the more notable because the one example they include with their account is *doesn't want to sound in the least like* (2006, 538).

8.3.5: The Repeated Comparative Construction

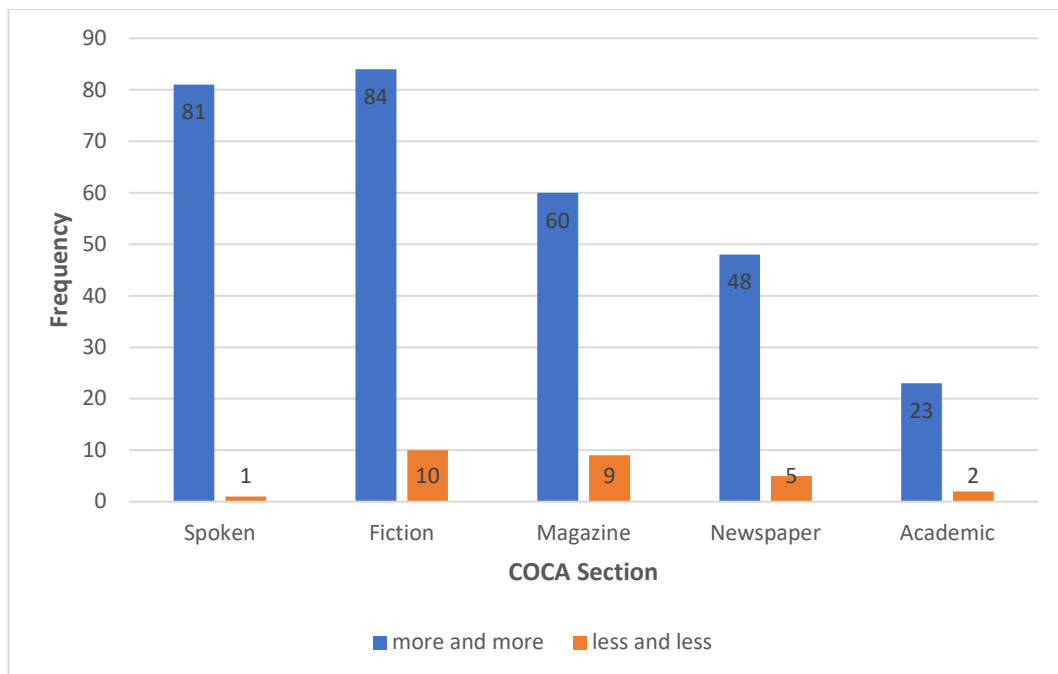
In their description of comparative and superlative forms, D&L describe the repeated comparative (*–er and –er*) construction which, they state, “is used to express a gradually increasing degree of the quality being described” (2006, 487). This represents one of the few times where D&L discuss a phraseological item, which in this case would be considered to be a collocational framework (Renouf and Sinclair 1991). Occurring 323 times, it is relatively frequently used with *like* in the COCA. The preference for using the superior comparative (*more*) over the inferior (*less*) is carried over to this construction as well. There are 296 occurrences of *more and more like* (as in Example 8.22) and only 27 occurrences of *less and less like* (as in Example 8.23). Similar to my findings for *less like*, the *less and less like* phrases are sometimes paired with *more and more like* (as in Example 8.24).

- (8.22) She's acting **more and more like** a candidate each day. (SPOK[ABC_ThisWeek])
- (8.23) The hobbled Green Bay Packers look **less and less like** a contender without injured Aaron Rodgers,

- falling to 4–3 on Sunday with a home loss to New Orleans. (NEWS[MinneapolisStarTribune])
- (8.24) He'd do anything for her, and even more for his namesake, the namesake that every day was looking **less and less like** the good doctor and **more and more like** one of the men Frieda used to know. (FIC[Bk: EleventhCommandment])

Figure 8.6 shows the frequency of usage for the repeated comparative constructions used with *like* in all five sections of the COCA.

Figure 8.6: Frequency of the repeated comparative construction in each section of the COCA, showing separate frequencies for *more and more like* and for *less and less like*



These data show that the repeated comparative *more and more like* is frequently used in all five sections of the COCA, with the most occurrences in spoken and fiction, and the fewest occurrences in academic writing.

8.4: Constructions Without Prescribed Rules

8.4.1: *As Much Like*

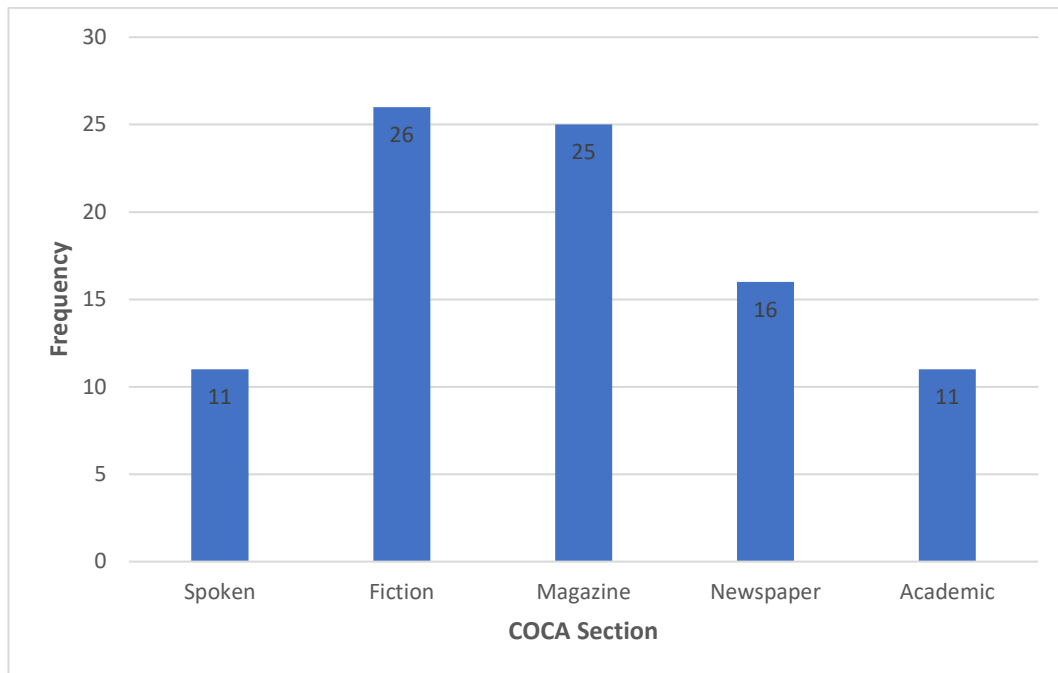
In their grammatical framework, D&L (486) stated that the *as ADJECTIVE* construction is used to show equality; however, there is one detail they did not mention in their brief analysis of gradable premodifiers used with prepositions. This construction does not work with prepositions, but the related construction, *as much*, was found to work with *like* and other prepositions. There are 98 occurrences of *as much like* found in the COCA, and Examples 8.25–8.27 show how this construction typically works with *like* to state equality between two noun phrases, in what appear to be mostly compositional statements.

- (8.25) Cookbooks from the country's best chefs look **as much like** art books as instruction manuals. (NEWS[WashPost])
- (8.26) He makes practice **as much like** a real game as possible. (SPOK[CBS_Sixty])
- (8.27) Make the learning situation **as much like** an authentic or real-world situation as possible to inspire and engage students. (ACAD[TeachLibrar])

Just over ten percent of the usage of *as much like* in the COCA is part of the lexicalized phrase *about as much like*, and this phrase will be discussed separately in 8.4.3.3.

Figure 8.7 shows the frequency of the *as much like* construction, across the five sections of the COCA.

Figure 8.7: Frequency of *as much like*, as found in each section of the COCA



These data show that *as much like* is used across all sections of the COCA, and that the most frequent usage is found in fiction and magazines.

8.4.2: *Enough Like*

Another item from D&L's (486) grammatical summary states that sufficiency can be expressed with the *ADJECTIVE enough* construction. This does not apply to prepositions, however. While there are 98 occurrences of *like enough* in the COCA, they do not function as comparative statements, and this can be seen in Examples 8.28–8.30.

- (8.28) It doesn't sound **like enough** oomph for a sports car, but slide behind the BRZ's wheel, down into the bolstered seats, and push the pedal. (MAG[PopMech])
- (8.29) In one last dash, she cooked up what seemed **like enough** curry paste to fill our freezer for several generations. (NEWS[CSMonitor])
- (8.30) Five minutes has dwindled into nothing, fifteen minutes barely feels **like enough** time to take a full breath, while the four hours of standing take an eternity. (FIC[Triquarterly])

In the above examples, we can see that *enough* is not combined with *like* to form a comparative structure, but rather it functions as a determiner for the noun phrase that follows. However, the reverse construction, *like enough*, does function as a comparative structure of *like*, and this is shown in Examples 8.31–8.33 below.

- (8.31) “The man he shot might look **enough like** me to pass,” he said, “especially with his face in a pillow and two bullets in his head. (FIC[Bk:HitList])
- (8.32) Because it had sounded **enough like** him that I could not say to myself or to Linda that this is not Ted’s writing. (SPOK[CBS_Sixty])
- (8.33) According to Andrew, dating on Twitter is **enough like** dating in the real world that it’s not much of a question. (MAG[TheVerge])

Figure 8.8 shows the frequency of *enough like* for samples taken from each section of the COCA.

Figure 8.8: Frequency of *enough like* in each section of the COCA

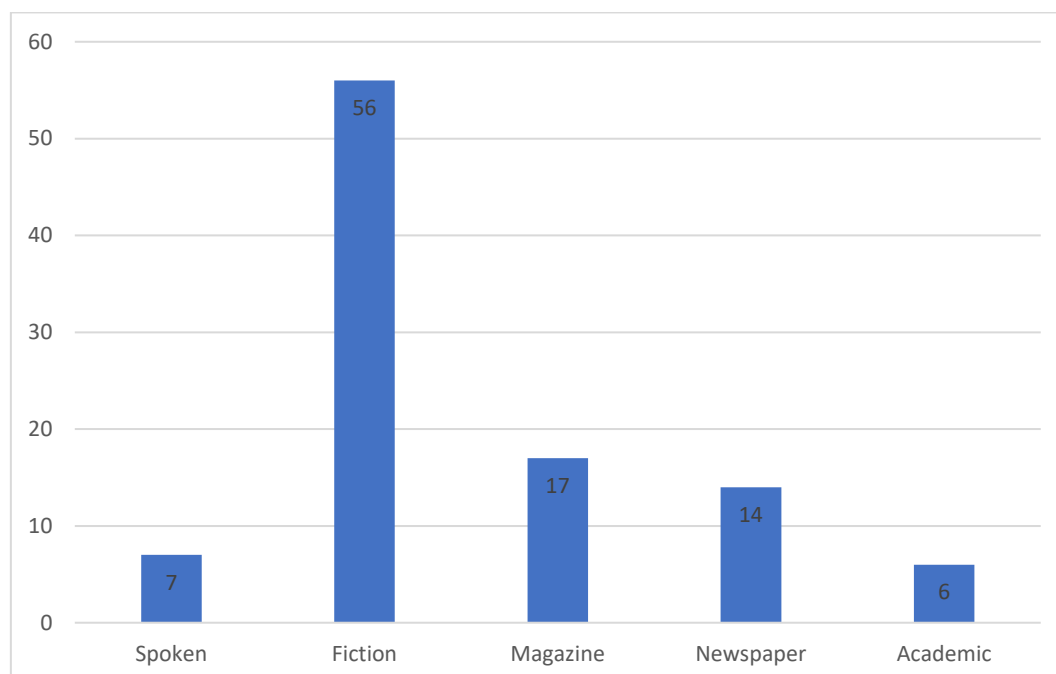


Figure 8.8 shows that the *enough like* construction is used most frequently in fiction, and this is true by a considerably large margin.

8.4.3: Grammatically Based Lexical Phrases

8.4.3.1: *BE More Like It*

While *more like it* is sometimes found in standard comparative statements, the most frequent usage of this construction is found in the semi-preconstructed phrase *BE more like it*. This usage is highly specialized, and it seems to perform various pragmatic functions; it could possibly be viewed as an example of attitudinal formulae, as found within the category of communicative phrasemes in Granger and Paquot's phraseological spectrum (see Figure 2.1). It was found in 254 out of the 461 total occurrences of *more like it* in the COCA. Examples 8.34–8.36 demonstrate this usage of *BE more like it*.

- (8.34) Dump is too clean a word, dive is too nice,
abomination **is more like it**.
 (FIC[Mov: FreddysDead])
- (8.35) We are not at all certain that consumerism is
 the driving force, as Lasch suggests. Paying the
 mortgage **may be more like it**. (NEWS[Houston])
- (8.36) I was very intimidated coming, and I came here
 to learn something new everyday, and I learn
 something everyday. Every minute of the day **is
 more like it**. (SPOK[CBS_48Hours])

In Example 8.34, the speaker is searching for the best term to describe a tavern that he has walked into, rejecting the first two nouns (*dump* and *dive*) before settling on the word *abomination*. The usage of *BE more like it* in this instance signals that the word finally chosen is better, or more appropriate, than the previous mentioned words. While Example 8.35 is similar, in that it involves signalling a better choice of words with *BE more like it*, by which the speaker is suggesting a correction or contradiction to the previous stated

theory. In Example 8.36, *BE more like it* is used for added emphasis on the intensity of a training program, comparing the program to his regular life. The construction is being used to highlight a change. While these three examples all demonstrate different pragmatic uses, they all share the function of stating an “improved” description of something stated previously. Often presented as part of a stated opinion, its meaning has been shown to belong as a common feature of *BE like* similes in Chapter 5.

Additionally, this usage of *BE more like it* often consists of creative word play by the speaker or writer, in which the “improved” description is made to sound very similar in form to the entity it is being compared to. The form is often borrowed from (or imported in response to) an interlocutor. This feature, highlighted in Examples 8.37–8.39, seems to be often used for dramatic or comic effect.

(8.37) I lied to myself and to the world, pretending I was doing public service. Public disservice **is more like it.** (FIC[Bk: WhenFiveMoons])

(8.38) Friendly skies? Fiendish skies **is more like it.** (MAG[Money])

(8.39) Clinton as commander in chief? Contradiction in chief **is more like it.** (NEWS[FoxNews])

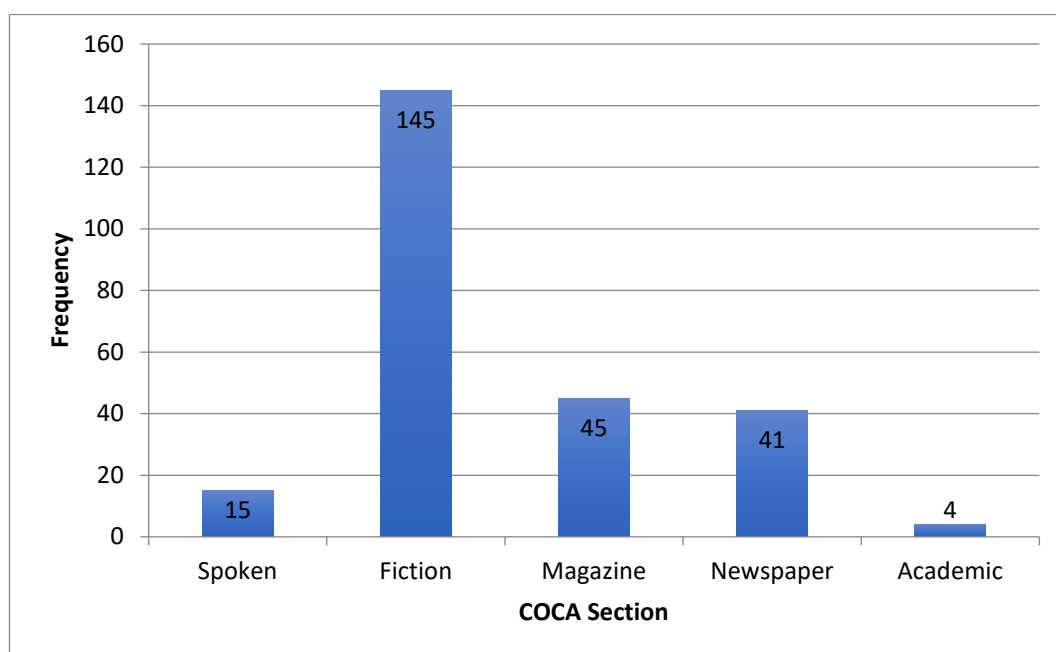
Attesting to the creative nature of this usage of *more like*, Examples 40–44 demonstrate how almost any part of speech can be used with *BE more like it*. Nouns and noun phrases are most commonly used in this phrase (as in Example 8.40), followed by verbs and verb phrases (as in Example 8.41) and adjectives, accounting for 45, 24, and 23 percent of occurrences in the COCA respectively. Also used, but much less frequently, are prepositional phrases (as in Example 8.42), adverbs (8.43), and full clauses (8.44), accounting for four, two, and two percent respectively. It

seems clear that most cases of *BE more like it* involve contrastive stress for emphasis, with the word or syllable that is being contrasted before *BE more like it* most likely to be stressed.

- (8.40) "Got the butterflies?" "Bats **is more like it.**"
(FIC[Bk:KillingChe])
- (8.41) "Sorry. Just thinking." Being rude **was more like it.** (FIC[Bk:PandorasBox])
- (8.42) Under control, my ass. Under indictment **is more like it.** (FIC[Mov: Crow3])
- (8.43) Once or twice, hell—twice a week **is more like it,**" Sheila said. (FIC[Atlantic])
- (8.44) You talked to her? She talked to me, **is more like it.** (FIC[SouthernRev])

The frequency data for *BE more like it* from each section of the COCA is shown in Figure 8.9.

Figure 8.9: Frequency of *BE more like it* in each section of the COCA



While this usage of *BE more like it* is found in all five sections of the COCA and follows the same order of frequency as *more like* in general, Table 8.9 highlights that the vast majority of this usage is found in fiction. This is not

surprising, as it fits with results discussed in Chapters 5, 6, and 7, in which phrases composed with *like* were most frequent in fiction. As Cowie (1998) put it,

the work of fiction is in fact the richest field for the deployment of phraseology for stylistic effect, since the literary author has access to the entire wealth of the language, and can draw on its expressive resources on various levels. (12)

Furthermore, the very low occurrence of this usage in the academic section suggests that this function of *more like* is mostly used in casual English. Of the four occurrences found in the academic section, two are presented as opinions such as with Example 8.45, and the remaining two are paraphrases of reported speech (see Example 8.46).

(8.45) the slogan “Opportunity Returns” is a rather coy understatement. “Boom” **is more like it**: the world is in the midst of an unprecedented love affair with coal. (ACAD[NaturalHist])

(8.46) “when I used the language of choice to her, Doris Paul immediately responded, ‘Choices, choices. Decisions, decisions would **be more like it.**’” (ACAD[AnthropolQ])

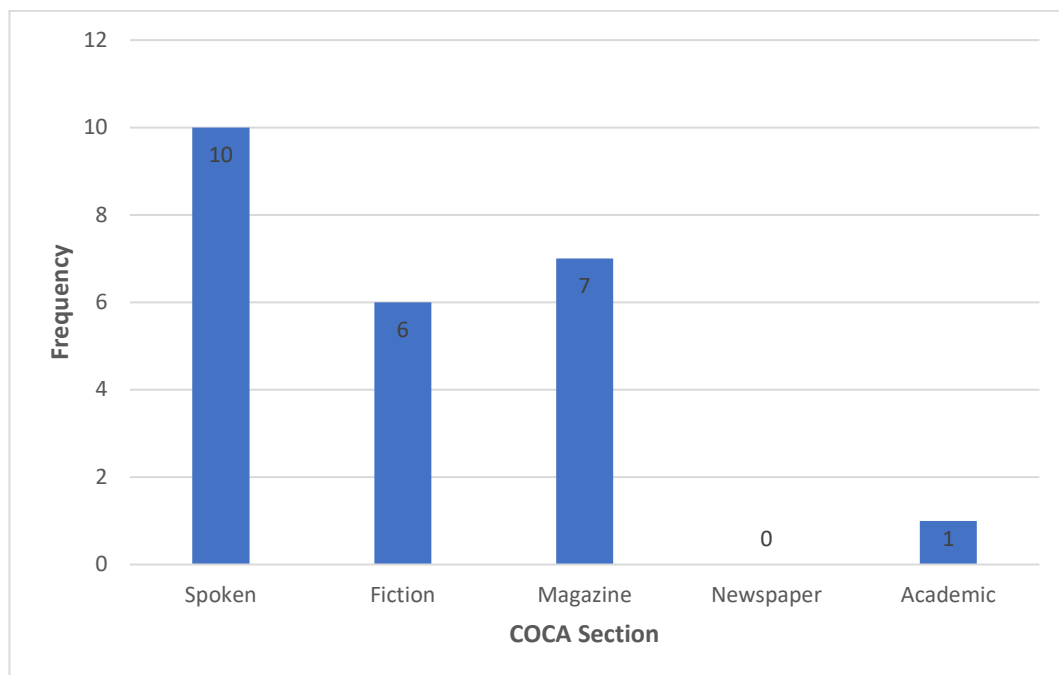
Now, this usage of *BE like* is still technically a comparative construction and does not specifically deviate from the grammatical description of comparative forms as described by D&L (2006, 486). However, it does represent a lexicalized, pragmatic usage of a comparative. This phrase is relatively frequent in the COCA, and while it appears to be a comparative form, it is much more complex than a simple comparison, and a grammar such as D&L’s is not capable of fully accounting for it. This leaves a gap in the description of the language.

8.4.3.2: *At Least Like*

At least like is a comparative form that is not included in the comparative/superlative summary provided by D&L (2006). It is relatively infrequent, occurring only 24 times in the COCA. However, it is used to indicate the minimum that is acceptable, and it is also used with the verbal form of *like*. *At least* occurs most frequently in the COCA with numbers and percentages. The data shows that there is no usage of *at most*.

- (8.47) Looks **at least** like you have a sunny day down there. (SPOK[ABC])
- (8.48) That is why I had tried to make this work, this time with Tusker and all that he could not give me, why I had tried to hide my head if not like an ostrich in the sand, then **at least** like a Canada Goose in the blueberry patches. (FIC[VirginiaQRev])
- (8.49) And prepare to live, if not like a king or queen, then **at least** like a Rockefeller. (MAG[Money])

Figure 8.10: Frequency of *at least like* in each section of the COCA



8.4.3.3: *About As Much Like*

There are 11 instances of *about as much like* found in the data. Only one of these (shown in Example 8.50) appears to be compositional in nature. The remaining ten occurrences the construction being used as a lexicalized phrase. This lexicalized form is used to signal an analogy, to express that one thing is not at all like another thing. See Examples 8.50–8.53.

- (8.50) You'll find both playability and control in the Adams Red. It's a true player's club that looks and handles **about as much like** an iron as any hybrid you'll find. (MAG[GolfMag])
- (8.51) Monden Kogyoku's 1975 "Bird" looks **about as much like** a bird as it does like a serviceable flower-container. (NEWS[SanFranChron])
- (8.52) Telling the difference between the track of an average buck and a doe can be difficult, but the track of a really big buck—one that you can stick your fist into—is **about as much like** the dainty imprint of a doe as a bobcat's is like one made by a Siamese. (MAG[FieldStream])
- (8.53) Oman is **about as much like** that part of the Middle East that has been making the evening news since biblical times with its internecine struggles as New York is like, well, New Mexico. (NEWS[Chicago])

8.5: Lexicalized Phrases Without Prescribed Rules

8.5.1: *That/This BE More Like It*

The second phrase found in the COCA that consists of *more like* while not fitting D&L's prescription for comparatives is *that/this BE more like it*. This phrase is less frequent than *BE more like it*, occurring 129 times in the COCA. While on the surface it also appears to be a straightforward comparative usage, closer inspection of the corpus data show that the pronoun *it*

does not usually have an inferable item of comparison. That item of comparison, according to D&L (2006, 487), is expected if the comparison is to be understood. In the corpus, it is used more often as a statement of contentment than as a comparison. Several examples of this specialized usage of *more like* are provided in Examples 8.54–8.56.

- (8.54) Now the stove lids were glowing. **"That's more like it,** eh? Uncle Labe said, shedding his coat. (MAG[BoysLife])
- (8.55) Using a belt sander, she was at last making progress on the job. Under the power tool's howling attack, the paint came off in clouds of sawdust. And **that was more like it.** (FIC[Bk: KnockdownHomeRepair])
- (8.56) They climb the gradient of the hill opposite without slowing, then disappear from sight in the broken terrain. **"That's a little more like it:** a salutary, bracing antidote to that tiresome baby. (NEWS[CSMonitor])

The phrase is used in both present and past tense, as shown in 8.54 and 8.55 respectively, and it can be used as a premodifier for hedging, as seen in Example 8.56, although this was found to occur rather infrequently.

Additionally, *that/this BE more like it* is often preceded by the adverb *now* (as in Example 8.57), or by the exclamation *ah* (as in Example 8.28).

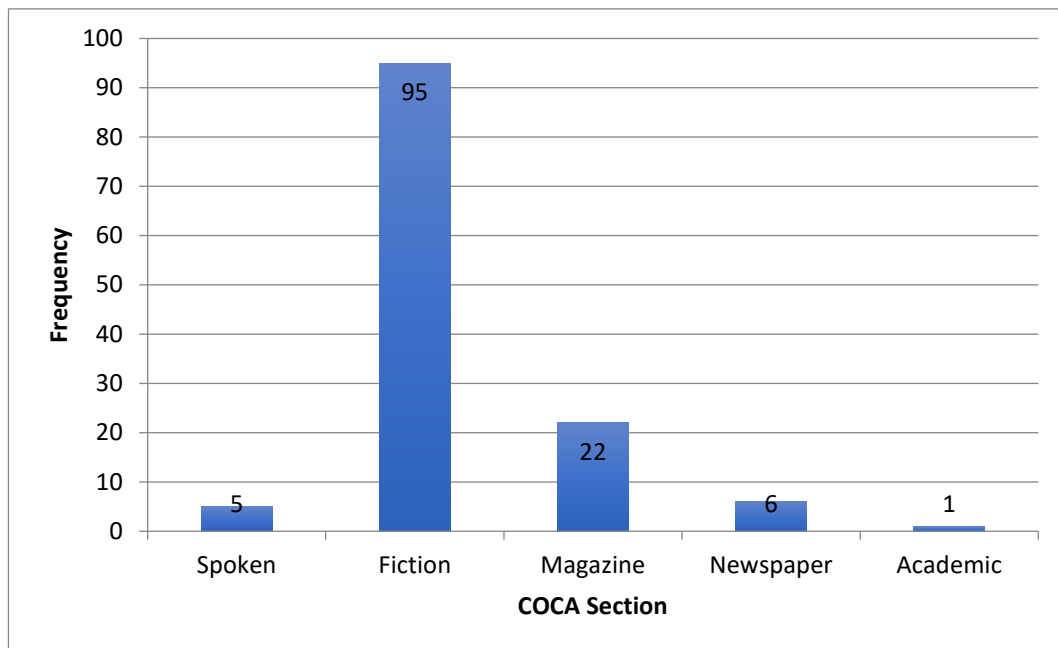
- (8.57) Erica makes it look so easy. Mr-Li: **Now that's more like it.** That was awesome, Erica! Good job, Erica. (SPOK[CBS_SunMorn])
- (8.58) Plate bindings hold the hard boots in a rigid coupling to the board. **"Ah, now this is more like it,** "I thought to myself the first day I tried the new system. (MAG[Skiing])

Three occurrences of *that/this SOUND more like it* (Examples 8.59–8.61) were identified in the data, and I examined these to determine if they were being used as a variation of the same idiomatic usage of *more like*. All

three occurrences represent spoken English; the first two are from fiction but belong to characters' utterances in dialogue. Examples 8.59 and 8.60 appear to belong to more of a standard comparative usage, where *that SOUND more like it* is being used to confirm an improved statement. Example 8.61, however, seems to sit in between the true comparative and the lexicalized phrase as it takes on the function of describing contentment. It is notable that a previous entity can be inferred as available for comparison. In this case, it is a less desirable acting role.

- (8.59) "You also say that sometimes the most crucial clue is not what has happened, but what has not." **That sounds more like it,**" I said.
(FIC[FantasySciFi])
- (8.60) There are a few chamalian males of any breed who can resist the secuction of a fate greater than their own." **"That sounds more like it.** It has to be powerful enough to sustain in the face of figures like these," Jerome said, motioning to the notes on his mindpad." (FIC[Analog])
- (8.61) And then finally, they came out and they told me, they said, Well, said, you'll be playing the role of a chief grave digger. I said, **That sound more like it,** I said, because, you see, when I was a kid, I used to dig and shovel coal out of the coal cars . . . [unintelligible] at night to keep warm. (SPOK[NPR_Sunday])

Figure 8.11 shows the frequency distribution of the idiomatic *that/this BE more like it* for each section of the COCA.

Figure 8.11: Frequency of *that/this BE more like it* from each section of the COCA

These data show that this phrase is also predominantly used in fiction, although it is often presented as reported speech. This *more like* phrase is also least frequently found in academic writing, with only one occurrence, shown in Example 8.62 below. This seems to suggest that this idiomatic usage is most often used in casual English.

- (8.62) To counter this common tendency, Humanis Manifesto II states in its twelfth article, "For the first time in human history, no part of h mankind can be isolated from any other. Each person's future is in some way linked to all." **Ah, that is more like it!** Instead of encouraging racial and ethnic pride, which ay well lead us into nationalistic difficulties and prevent us from having one group rate itself as "good" and another as "bad", let us perhaps encourage pride in us all being human rather than being a member of a certain group. (ACAD[Humanist])

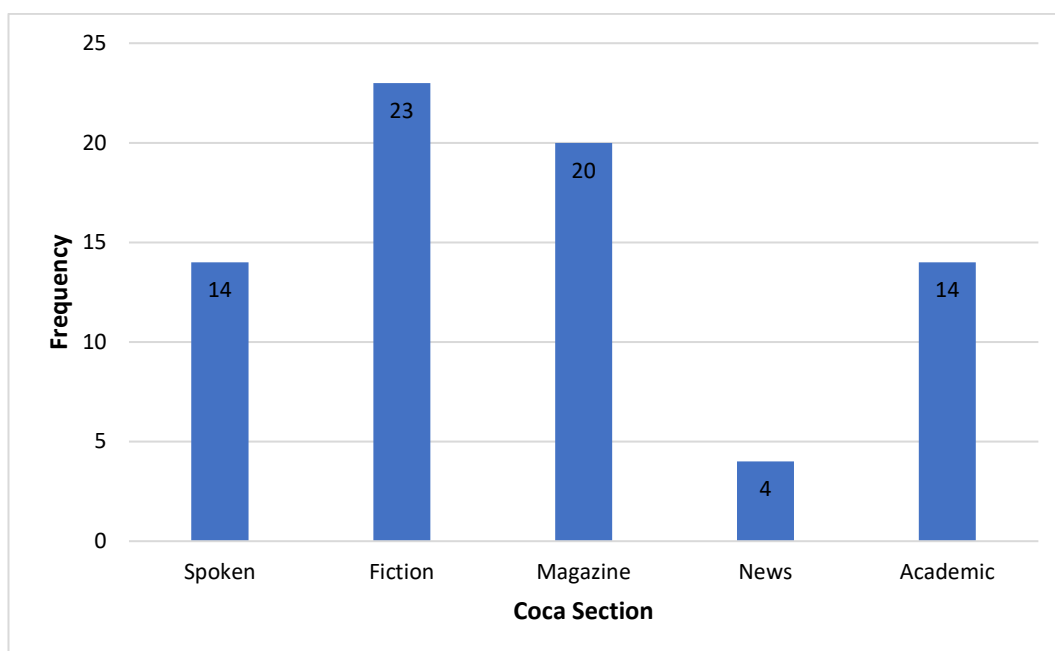
While *BE more like it*, as discussed in 8.2, behaves as a lexicalized construction, *that/this BE more like it* is more fully lexicalized. That is, it seems

completely removed from its origin as a comparative form because there is often no inferable complement for the comparison. This phrase provides further evidence for the inadequacy of traditional grammars which do not address lexicalized phrases such as these.

8.5.2: *More or Less Like*

The phrase *more or less like* occurs 72 times in the COCA. Rather than having a comparative function, it takes the meaning of *similar to*. This lexicalized usage, illustrated in Examples 8.63–8.65, is not mentioned in D&L’s grammatical explanation.

- (8.63) The United States trying to catch up with
Britain behaved **more or less like** the leaders of
Meiji (and postwar) Japan trying to catch up
with the United States. (MAG[Atlantic])
- (8.61) It was **more or less like** gutting a catfish,
except that the cut was horizontal rather than
vertical. (NEWS[Houston])
- (8.62) The collected data were fed into an avatar-like
device, which, on that groundwork, as an effect
“behaved” to a notable degree like a human
actor, and **more or less like** its original—Bina
Aspen. (ACAD[ReviewOfContemporaryPhilosophy])

Figure 8.12: Frequency of *more or less like* in the COCA by section

8.6: The Importance of Phraseology

In this chapter, I have explored how prepositional *like* behaves when paired with comparative and superlative premodifiers. By comparing the corpus data with a traditional grammatical explanation of comparative and superlative forms, I have observed two main findings. First, considering the vast complexity of a language, even comprehensive traditional grammars such as D&L's fail to cover the full spectrum of grammatical items. This is evident in the number of grammatically based items found in the corpus data that were not included in their grammatical framework. Second, as demonstrated in this analysis, traditional grammars seldom take phraseology into consideration and, because of this, frequently used lexical phrases are not accounted for. This can be detrimental to language learners.

9

Discussion

Pedagogic Implications and Applications

9.1: Comparing the COCA and the ELTCC

In this chapter, I investigate the treatment of *LIKE* in ELT coursebooks by reporting on a comparative analysis of findings from the Corpus of Contemporary American English (COCA) and the English Language Teaching Coursebook Corpus (ELTCC). The latter is a specialized corpus of general communicative ELT coursebooks compiled for the current study (see Chapter 4 for an in-depth discussion of the corpora used in this study). This comparative analysis will provide the relevant data necessary to carry out an informed discussion of the pedagogic implications and applications of the current study. First, in section 9.2, I provide an overview of the current state and main trends regarding the treatment of phraseology in ELT coursebook design. Next, in section 9.3, I compare and discuss the frequency distributions of the main senses and functions of *LIKE* in the ELTCC and in the COCA. In section 9.4, I examine the most frequent patterns of usage of *like* in the ELTCC and compare this with the corresponding data in the COCA. Finally, in section 9.5, I discuss the pedagogic implications and applications of these findings and make some recommendations on how coursebook designers might be able to incorporate the corpus-based pedagogic applications of the lexical syllabus and data-driven learning into future

coursebooks to better deal with vocabulary in general, and high-frequency words such as *LIKE* in particular.

9.2: Corpus Linguistics and ELT Coursebooks

It has been well established in the literature that ELT teaching materials and classroom practice often do not reflect current applied linguistics research findings (Burton 2012). In his investigation of the influence of applied linguistics on ELT coursebooks, Littlejohn (1992, 117) found “little evidence of a very strong link between applied linguistic discussion and the design of [ELT coursebooks].” Meanwhile, Chalker (1994, 41) found few differences in content and organization between coursebooks from the 1950s and 1960s (when structuralism was the dominant pedagogical approach) and others from the 1990s. By the 1990s, communicative language teaching had become established as the new dominant approach in ELT—but perhaps not in publishing. More recent studies have confirmed that little has changed with regard to the influence of research in applied linguistics (and corpus linguistics in particular) on ELT practices. O’Keeffe et al. (2007) note that there is a “frequent mismatch” between corpus linguistics research, teaching materials, and classroom practice, stating that

widespread use of “corpus linguistics” does not mean that the term or its findings are necessarily fully or widely understood in the context of language pedagogy. In addition, many important developments in the field of corpus linguistics are not always communicated or usefully mediated in terms of their implications for language teaching. (xi)

This mismatch between research and practice is evident in the treatment of phraseology evident in contemporary ELT coursebooks. A study on the vocabulary selection process by Koprowski (2005) examined the lexical syllabuses of three coursebooks by assigning usefulness scores for each

multi-word lexical item. This was based on frequency and range (the number of text types an item is found in) using the 330-million-word Bank of English corpus. Koprowski concluded that nearly a quarter of the multi-word lexical items included in contemporary ELT coursebooks were “of limited pedagogical value to learners” (322) as they occur very infrequently and in limited contexts. Koprowski did acknowledge that commercial textbook publishers seem to be aware of the importance of multi-word lexical items, based on their abundant inclusion in coursebooks. However, he warned that the selection process was highly subjective and often conducted without reference to corpus data, amounting to “an unprincipled and careless selection process” (328). In another related study, Li (2015) conducted a corpus-based study of the highly-frequent nouns *time* and *thing* and concluded that the selection and presentation of phraseology in ELT coursebooks used in China is also problematic.

Building on the work of Koprowski (2005) and Li (2015), I conducted a study (Peppard 2016) and found that the “unprincipled and careless selection process” of multi-word items described by Koprowski applies to single-word items as well. The lexical items included in a communication-based coursebook I was using in my first-year General English university course, *Nice Talking with You 2* (NTWY2; Kenney 2012) at times appeared rather questionable. The vocabulary list for the unit entitled “My Place,” which focused on describing your house or apartment, and your hometown, consisted of 29 items. Only six of the 29 items in this unit consisted of more than one word, and one of the single items included was the word *yuppie*. This word appears only twice in the entire ELTCC, and both of these occurrences are in NTWY2. One was in the vocabulary list, and the other was in

a corresponding gap-fill exercise to complete the sentence *I live near a lot of tech companies, so a lot of my neighbors are yuppies* (Kenny, 2012).

A search of the then-520-million-word COCA confirmed that *yuppie/s* was rather infrequent, occurring 1.75 times per million words, and that its usage has steadily declined since 1990, from 3.38 words per million to 0.58 words per million. In contrast, a seemingly useful word for this topic, the noun *rent*, was not included in the *NTWY2* vocabulary list, yet was found to occur 6.84 times per million words and also had a stronger range than *yuppie/s* (Peppard 2016, 24). This finding does seem to suggest a ‘subjective and careless’ vocabulary selection process for this coursebook.

Furthermore, a study conducted by Shortall (2007) showed that the mismatch between corpus linguistics research and ELT coursebook design extends to grammatical description as well. Comparing the present perfect (PP) as used in the BOE’s 20-million-word spoken British corpus (brspok) with the PP featured in 32 textbooks ranging from beginner to intermediate level, Shortall found that textbooks tended to unnaturally isolate PP. Corpus evidence shows that PP often interacts with passive voice and modal verbs. In addition, it was found that the textbooks misrepresent the frequencies of functional adverbials with PP, giving priority to *ever*, *yet* and *since* rather than *now*. Emphasizing *now* was usually reserved for demonstrating the present continuous, although it appears frequently in the corpus PP data.

The following sections, by examining the treatment of *like* in ELT coursebooks through an analysis of the ELTCC and comparing these findings with corresponding data from the COCA, aim to build on the research outlined above. These results will then be used to further explore the perceived mismatch between ELT coursebooks and natural language use, and

to determine if there is evidence of “subjective and careless” treatment of the high-frequency word *like*.

9.3: Frequency of *LIKE* in the ELTCC

Comparing the frequency data for *like* in the ELTCC and COCA reveals significant differences between the two corpora. While *like* is highly frequent in both the ELTCC and the COCA, it is much more so in the ELTCC, where it is ranked as the 34th most frequent flemma. This ranking includes both prepositional and verbal usage of *LIKE*, as the ELTCC does not feature part-of-speech (POS) tagging (see McClean (2018) for a discussion of the flemma and its pedagogical application). In the COCA, which is much larger and more advanced with POS tagging, we can see that as a preposition, *like* is the 71st most frequent word. As a verb, it is the 144th most frequent word, while as a conjunction it is the 689th most frequent word. *LIKE* can also be used as an adverb, adjective, and conjunction, although much less frequently—with rankings of 1157th, 1234th, and 4868th respectively. Table 9.1 provides the frequency and per-million data for each word form of *LIKE* in both corpora.

Table 9.1: Frequency and per-million distributions of *like* word forms in the English Language Teaching Coursebook Corpus (ELTCC) and in the Corpus of Contemporary American English (COCA)

<i>LIKE</i> Form	ELTCC Frequency	Per million	COCA Frequency	Per million
like	3,774	3,667.3	1,240,789	2,027.4
likes	288	279.8	22,584	36.9
liked	78	75.8	36,834	60.2
liking	9	8.7	2,736	4.5
	4,149	4,031.7	1,302,943	2,129.0

The data in Table 9.1 show that all word forms of *LIKE* are more frequent in the ELTCC. The base form *like* is 1.8 times more frequent in the ELTCC than the COCA, at 3,667.3 and 2,027.4 words per million

respectively. The most noteworthy discrepancy between the two corpora, however, is with the conjugated form *likes*. It is 7.6 times more frequent in the ETLCC, with 279.8 and 36.9 words per million respectively. This large discrepancy is a reflection of a major difference in the presentation of *like* in the ELTCC with regard to function and sense when compared to the COCA. It is clear that use of the verbal form of *like* is much more prevalent in ELT coursebooks than in more general-purpose English-language texts.

Figure 9.1 and Table 9.2 below show the results from sorting 500-concordance-line samples of *LIKE* by function and sense for both the ELTCC and COCA, revealing highly contrasting frequency distributions between the two corpora.

Figure 9.1: Frequency distributions for the three main functions of *like* in the ELTCC and COCA

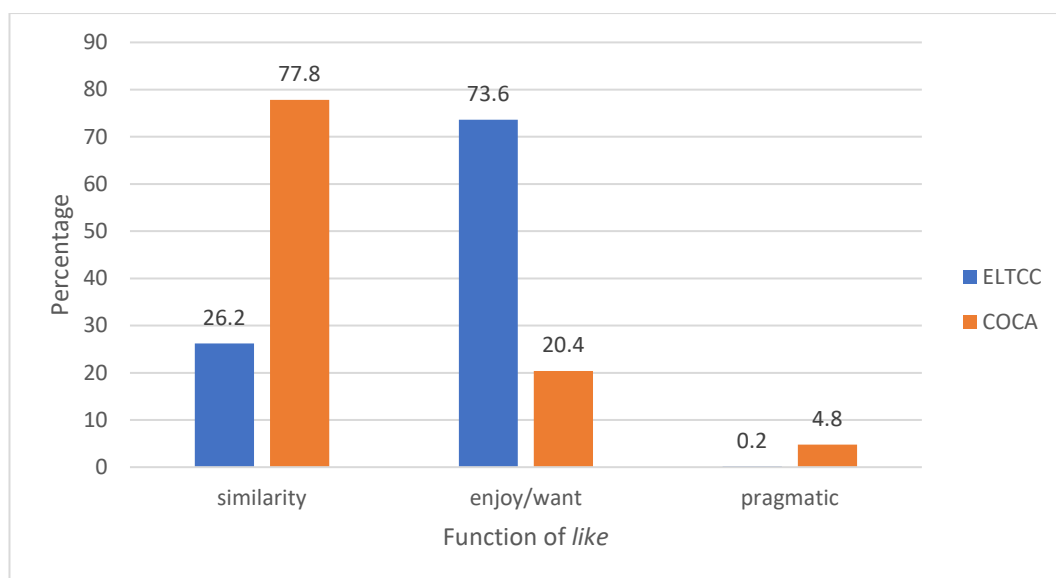


Figure 9.1 highlights the major discrepancy between the ELTCC and the COCA, with nearly opposite frequency distributions for the two most frequent functions of *like*. The majority of *like* usage in the ELTCC is concerned with enjoyment (*like*₂) and mostly verbal, accounting for 73.6 percent of all

occurrences in the sample. Only 26.2 percent of the sample is concerned with similarity (*like*₁), and the pragmatic usage of *like* (*like*₃) is negligible, with only one occurrence (0.2%). These data show that the treatment of *like* in ELT coursebooks does not reflect wider usage as reflected in the COCA, where the frequency distribution is nearly the opposite of that found in the ELTCC. The majority of *like* usage in the COCA is concerned with similarity (*like*₁), accounting for 73.6 percent of the sample, and only 20.4 percent of the sample is concerned with enjoyment (*like*₂). Also, the COCA has a much higher rate of pragmatic usage of *like* (*like*₃), accounting for 4.8 percent of the sample.

Table 9.2: Frequency and percentage distributions for the functions and senses of *like* in 500-concordance-line samples from the ELTCC and COCA

Function/sense	ELTCC Frequency	Percent	COCA Frequency	Percent
<i>like</i> ₁ similarity	131	26.2%	374	74.8%
1.1 similar to	108	21.6%	302	60.4%
1.2 same as	3	0.6%	36	7.2%
1.3 for example	20	4%	36	7.2%
<i>like</i> ₂ enjoy/want	368	73.6%	102	20.4%
2.1 fond of	270	54%	74	15%
2.2 want	97	19.4%	26	5%
2.3 in the mood for (FEEL like)	1	0.2%	2	0.4%
<i>like</i> ₃ pragmatic	1	0.2%	24	4.8%
3.1 filler	1	0.2%	17	3.4%
3.2 discourse marker	0	0%	7	1.4%

Table 9.2 provides a more detailed breakdown of the corpus data, one including the frequency distributions for the main senses of each function of *like*. Here, we can see that despite the major difference in functional usage, the proportion of sense usage within each function is similar for both the ELTCC and the COCA—with only two exceptions. First, the usage of *like*_{1.3} (*for example*) is notably more frequent in the ELTCC. A closer look at the concordance data, reveals the reason for this frequent usage of *like*_{1.3} in the ELTCC is due to instructional language found in language production

activities commonly used in coursebooks (see 9.4.2.2). Second, pragmatic usage of *like* (*like*₃) is highly infrequent in the ELTCC, with only one occurrence of *like* being used as a filler, and no occurrences of *like* being used as a discourse marker (see 9.4.3). These conflicting results will be examined and discussed in more detail in section 9.4 below, with comparative analyses of the most frequent *like* constructions found in the two corpora.

9.4: *LIKE* Constructions in the ELTCC

In this section, the usage of *like* in the ELTCC will be examined and compared with the corresponding data in the COCA. Results will be presented from the most frequent to least frequent functions and senses of *like* usage in the ELTCC.

9.4.1: *Like*₂ (*Enjoy/Want*)

9.4.1.1: *Like*_{2.1} (*Fond Of*)

As briefly mentioned in Section 9.3 (and shown in Table 9.2) above, the vast majority of occurrences of *like* in the ELTCC sample (73.6%) involve the verbal form of *like* used for showing fondness for or wanting something (*like*₂). The most frequently used sense of *like* in this form (*like*_{2.1}), used to show fondness for something, accounts for more than half (54%) of the sample. Comparing the treatment of *like*_{2.1} in the ELTCC with its treatment in the COCA reveals several notable discrepancies.

First, sentences in the ELTCC tend to be shorter than those in the COCA. Of the 270 occurrences of *like*_{2.1} in the sample, 127 (47%) consist of simple declarative statements—mostly in the singular first person, as shown in Examples 9.1–9.3.

(9.1) I **like** listening to music. (CC[EngFirstHand1])

(9.2) I **like** to read before going to sleep.
(CC[Passages1])

(9.3) I **don't like** the movie. (CC[AmEngFile1])

The sentences in the ELTCC have an average length of 7.5 words, whereas corresponding sentences in the COCA that show declarative *like*_{2.1} tend to be nearly twice as long. With an average of 14.5 words, they consist of compounds and more complex clause structures, as shown in Examples 9.4–9.6 below. This discrepancy is understandable, due to the fact that language is often simplified when produced for the purpose of language teaching.

(9.4) I **like** the fact that you're using the whole game --kicking off, punting and doing all that—but I'd like to see both teams have the chance to have the ball. (News[Chicago])

(9.5) If they **don't like** the president, that's fine, quit. (Spok[CNN:AndersonCooper])

(9.6) But if I go someplace and it's supposed to be a good time, I **don't like** that it's political. (News[SanFranChron])

Wh- questions (as in Examples 9.7 and 9.8) and *yes/no* questions (as in Examples 9.9 and 9.10) with *like* can be found in nearly equal proportions in the ELTCC. They account for 14.3 percent and 11.3 percent, respectively, of the *like*_{2.1} occurrences in the sample.

(9.7) What do you **like** to do in your free time? (CC[WorldLink1])

(9.8) What kind of music do you **like**? (CC[NewInt1])

(9.9) Does Tyler **like** Ayumi's painting? (CC[WorldLink1])

(9.10) Do you **like** mornings? (CC[AmEngFileStart])

Compared to the ELTCC, questions with *like*_{2.1} are much less common in the COCA, with only one *yes/no* question (as in Example 9.11) and no *wh-* questions among the 102 occurrences of *like*_{2.1} usage in the sample.

(9.11) "Why do you **like** me?" Rita said. (FIC[MassachRev])

The higher proportion of questions in the ELTCC is again due to the pedagogic orientation of coursebooks. Many questions are directed at the language learner in order to provide opportunities for language production and, based on comparison to the COCA, it appears that this mode of questioning is not as high in everyday language use.

Another feature of *like*_{2.1} usage in the ELTCC, different from the more natural language use reflected in the COCA, is the higher proportion of imperative sentences. There are 12 occurrences of imperative sentences in the sample, accounting for 4.4 percent of *like*_{2.1} usage. These sentences function to give instructions for student activities in the coursebooks, as in the following examples:

- (9.12) Think of a sport or activity that you **like** to do. (CC[WorldLink1])
- (9.13) Name one place tourists **like** to visit in your city. (CC[WorldLink1])
- (9.14) Check (✓) the features you **like**. (CC[Touchstone2])

Imperative sentences using *like*_{2.1} appear to be used rather infrequently in everyday examples, as there are no occurrences of this usage in the sample from the COCA.

Other differences also exist between the two corpora, particularly with regard to premodification of the *like*_{2.1} constructions. While the proportion of premodification occurrences are similar, with 16 and 12 occurrences for the ELTCC and COCA respectively, there is a notable difference in the specific premodifiers used. The intensifier *really* is overrepresented in the ELTCC, accounting for 13 of the 16 premodified examples in the sample (see Example 9.15), including two occurrences of the *DO not really like* construction (as in Example 9.16).

- (9.15) Well, I **really like** soccer. (CC[Touchstone2])
- (9.16) No, we **don't really like** table tennis.

(CC[FourCorners2])

Premodification of *like*_{2.1} in the COCA is much more diverse than in the ELTCC, with only two premodifiers that occur with *like* more than once: *also* (see Example 9.17) and *always* (Example 9.18), with two occurrences each. There is only one occurrence of *really like* (Example 9.19) in the data sample from the COCA.

- (9.17) I **also like** the fact that I can cut a panel down after I begin painting if I decide it does not have the right proportions for the landscape. (MAG[AmArtist])
- (9.18) I had **always liked** to think I had healthy eating habits, but a lot of men like to believe that about their selves. (FIC[FanFic])
- (9.19) I think you have to think about what you **really like** to do. (NEWS[NYTimes])

While the vast majority of *like*_{2.1} usage, in both the ELTCC and the COCA, involves the verbal form of *LIKE*, the nominal phrase *likes and dislikes*, is disproportionately frequent in the ETTCC. This is predominantly used as a section header, and for giving instructions to students for speaking activities, as in the following examples:

- (9.20) Talking about **likes and dislikes**
(CC[WorldLinkInt])
- (9.21) Talk about your entertainment **likes and dislikes**. (CC[EngFirstHand1])
- (9.22) Share your **likes and dislikes** with other people like you. (CC[NiceTalkWithYou])

This phrase occurs ten times in the 500-concordance-line sample, and 66 times in the entire ELTCC, or 64.1 times per million words. In contrast, this phrase is rare in the COCA, occurring only 243 times in the 663 million-word corpus, or 0.4 times per million. It appears most often in the magazine section (0.68 per million) and least often in the spoken section (0.14 per million).

No instances were found in the COCA of this phrase being used as a section header similar to its usage in the ELTCC.

- (9.23) We all like places where they know our names and are familiar with our **likes and dislikes**.
(MAG[TownCountry])

A final point of interest with *like*_{2.1} usage in the ELTCC data sample is the occurrence of the more lexicalized phrase, *LIKE it or not*, which is used to show that someone has no choice with regard to some situation or event taking place.

- (9.24) They also said it would help prepare us for the future because whether we **like it or not**, dress codes are everywhere. (CC[ComStrat3])
(9.25) **Like it or not**, people judge us by how we dress and take care of our appearance. (CC[ComStrat3])
(9.26) **Like it or not**, many of us are sleeping less on average. (CC[Passages1])

This expression occurs only once in the 500-concordance-line sample, and a search of the entire ELTCC revealed a total of four occurrences (3.9 per million). Three of these four occurrences are found in a single listening script, from an activity in one of the coursebooks. Thus, the phrase occurs in only two of the 35 coursebooks in the corpus. This finding supports Koprowski's (2005, 328) claim that vocabulary selection is often highly subjective in ELT coursebooks.

LIKE it or not is much more common in the COCA, occurring a total of 1,467 times (2.39 times per million) across all five sections. Most often, it is part of the longer construction *whether PRON like it or not*, which only occurs once among the four ELTCC occurrences. It is most frequent in spoken English, occurring 377 times (or 3.39 times per million), and it is least common in academic English (occurring 122 times, or 1.19 times per million).

- (9.27) They will raise taxes **whether we like it or not**.
(SPOK[Fox_Susteren])
- (9.28) This is what I'm going to do **whether you like it or not**. (MAG[Essence])

Considering the relative frequency of this *like* construction in everyday language use, it is notable that it is found in only two of 34 ELT coursebooks.

9.4.1.2: *Like*_{2.2} (*Want*)

There are 97 occurrences of *like*_{2.2}, meaning *to want*, in the ELTCC sample, accounting for 19.4 percent of the total usage of *LIKE* and a total of 347 occurrences per million in the entire ELTCC. Of these, 141 occurrences show *would like* (see Examples 9.29 and 9.30) and 199 occurrences show the contracted —'d *like* form, as in Examples 9.31 and 9.32.

- (9.29) The woman **would like** to buy some earrings.
(CC[CutEdgePreInt])
- (9.30) My family **would like** me to spend less on clothes. (CC[Passages2])
- (9.31) I'd **like** to change this into pesos, please.
(CC[CutEdgePreInt])
- (9.32) I'd **like** a hamburger and a large order of french fries please. (CC[NewInt1])

As can be seen in these examples, this *like* construction is grammatical in structure, consisting of a modal (typically *would*) paired with *like* and complemented with an infinitive or noun phrase. Of the 340 total occurrences of *would like* functioning as *like*_{2.2} in full sentences, the vast majority (298) consist of the *would like to VERB* pattern. Only 33 of the 340 instances of *would like* occur in the *would like NOUN* pattern. It is typically presented in coursebooks as a polite structure for expressing wants, offers, and requests, and is often found in sections on ordering food. *Would like to* is often used with travel.

Occurrences of *like*_{2.2} are much less frequent in the COCA, with only 26 instances in the data sample, accounting for 5.2 percent of all *like* usage. In the full 612-million-word COCA the *would like* construction occurs 87.9 times per million (53,647 occurrences), one-quarter of the ELTCC frequency. Instances in the COCA are evenly split between *would like* and —'d *like*, with 27,047 and 26,647 occurrences respectively.

- (9.33) I am seeking legal action and **would like** a public retraction of his comments.
(MAG[Mashable])
- (9.34) He also said the sheriff **would like** to see him in his office. (FIC[Dialogue:Mormon])
- (9.35) I'd **like** a piano lesson, please.
(FIC[Movie:GroundhogDay])
- (9.36) What maneuvers have you been reading about that you'd **like** to try out? (MAG[PopMech])

As Examples 9.33–9.36 suggest, the grammatical objects of the *would like* sentences found in the COCA tend to be much more varied in the COCA than in the ELTCC. While food and travel are commonly found as objects in the ELTCC, only one example could be identified in a 500-concordance-line sample of *would like* from the COCA.

There are 15 occurrences of *would like* in the ELTCC in which the structure itself is being described or included in a list of vocabulary items.

- (9.37) I **would like** is more polite than I want.
(CC[CutEdgePreInt])
- (9.38) Gerund and infinitive: like, would like, enjoy, love, hate (CC[NiceTalkWithYou2])
- (9.39) Would like + infinitive (CC[FourCorners2])

These types of sentences are obviously used for pedagogic purposes and no similar examples were identified in the COCA. Interestingly, pedagogic sentences such as these are not found in the ELTCC for describing *like*_{2.1}. Perhaps this is due to the assumption that students are already familiar with

the verbal use of *like* as it is prominently featured throughout the coursebooks.

Regarding longer lexical patterns, the most frequent recurring *like*_{2.2} pattern in the ELTCC, *N would like to thank N*, is concerned with author and publisher dedications in the front matter of most coursebooks. This pattern (found in Example 9.40), along with several verbal variations (as in Examples 9.41 and 9.42), occurs 26 (25.3 per million) times in the corpus, accounting for 21.8 percent of all *N would like to V* occurrences.

- (9.40) I **would like to thank** my co-authors of this book, Tracy Davis and Suzanne Rizzo, for their great collaboration and enthusiastic work while teaching English at Shantou University.
(CC[ComStrat3])
- (9.41) The **authors would like to acknowledge** their debt to: Michael Swan's Practical English Usage and John Eastwood's Oxford Guide to English Grammar.
(CC[AmEngFile4])
- (9.42) The **authors would like to dedicate** this book to Krzysztof Dabrowski. (CC[AmEngFile3])

The *N would like to thank N* pattern is also relatively frequent in the COCA, with 1,032 total occurrences (not including verbal variations). However, when compared to the ELTCC, it is used much less commonly, and occurs only 1.7 times per million words across all five sections and most frequently in the academic section, at 5.9 occurrences per million words. While this phrase may be useful for students of academic writing, considering its purpose and location in coursebooks, students are not likely to notice this phrase.

Nearly half of the 97 *like*_{2.2} occurrences in the ELTCC data sample consist of questions, with 28 *yes/no* questions (as in Examples 9.43 and 9.44), and 15 *wh-* questions (seen in Examples 9.45 and 9.46).

- (9.43) **Would you like** to have an ice cream?

- (CC[CutEdgePreInt])
- (9.44) Is there anything else that you'd **like** to try?
(CC[AmEngFile4])
- (9.45) Where **would** Susan **like** to travel? (CC[NewInt1])
- (9.46) Which people **would** you **like** to meet?
(CC[EngFirstHand1])

Similar to the results of the *like*_{2.1} comparison, questions with *like*_{2.2} are much less common in the COCA, with only the following three occurrences in the data sample:

- (9.47) Ideally, how long **would** you **like** to serve in the House? (SPOK[NPR_ATC])
- (9.48) **Would** you **like** to play a song for us, Jonathan?
(SPOK[NPR_AskMe])
- (9.49) Or is there something else you'd **like** to add?
(FIC[BkSF:JaySilentBob])

9.4.1.3: *Like*_{2.3} (*In the Mood For*)

Only one instance—Example 9.50—of *like*_{2.3} (*FEEL like*) was identified in the 500-concordance-line sample from the ELTCC. A search of the entire corpus, however, found this function of *like* to be highly frequent, accounting for 47.6 percent (30 out of 63) of all *FEEL like* occurrences. The majority of these instances of *like*_{2.3} (22 out of 30) were found in the *FEEL like V-ing* pattern (see Example 9.50) and there were four instances of *FEEL like N* (Example 9.51). Meanwhile, the remaining instances were pedagogic in nature, including gap fills and language description (Example 9.52).

- (9.50) I always **feel like** going dancing!
(CC[Passages1])
- (9.51) Actually, I don't **feel like** a movie.
(CC[EngFirstHand1])
- (9.52) **Feel like** can also mean "want / would like,"
e.g., I don't **feel like** going out = I don't want
to go out. (CC[AmEngFile4])

Two occurrences of *like*_{2.3} were found in the corresponding 500-concordance-line sample of *like* from the COCA. However, a follow-up 500-line sample of *FEEL like* from the COCA produced far fewer occurrences of *like*_{2.3} (28 occurrences, 5.6%) compared to the ELTCC findings. Of those in the COCA, 20 were found in the *FEEL like V-ing* pattern, and eight were found in the *FEEL like N* pattern, similar to the proportions of usage in the ELTCC.

(9.53) Drive on whatever side of the road you **feel like** driving on. (MAG[PsychToday])

(9.54) “Which means the only reason I’m taking off the lipstick is because I **feel like** it.”
(FIC[VirginiaQRev])

Although *feel like*, used in this way, is common in everyday language, these results suggest that the construction is overrepresented in ELT course-books.

9.4.2: *Like*₁ (Similarity)

9.4.2.1: *Like*₁ (Similar To)

The BE like construction. Table 9.3 compares frequency and per-million distributions for all forms of the *BE like* construction, in both the ELTCC and the COCA.

Table 9.3: Frequency and per-million distributions for all forms of *BE Like* in the ELTCC and COCA

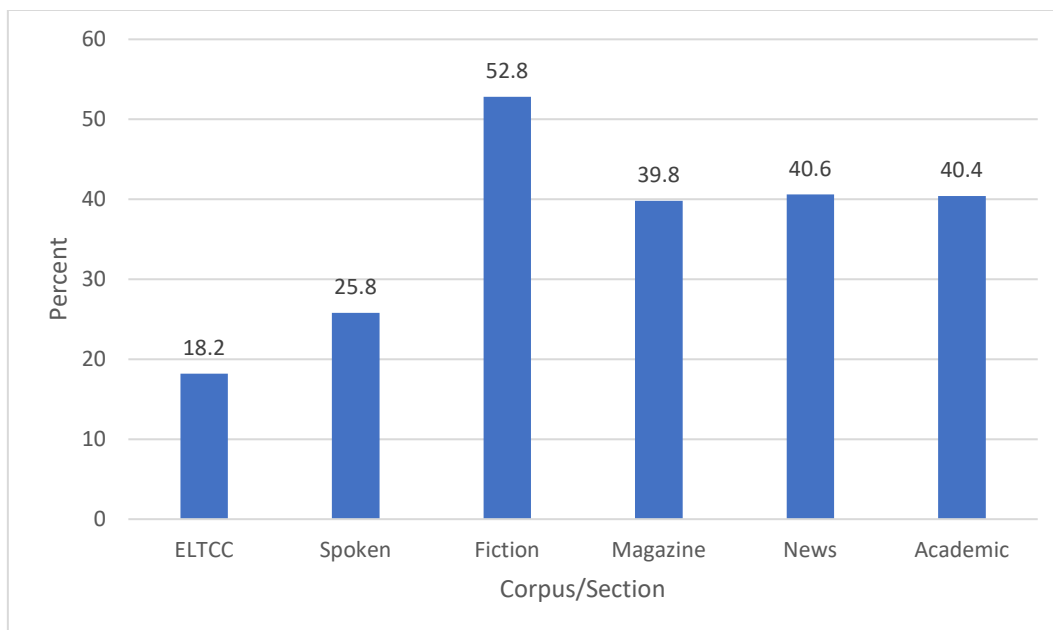
<i>BE like</i> form	ELTCC Frequency	Per million	COCA Frequency	Per million
<i>'s like</i>	10	9.7	31,993	52.3
<i>was like</i>	4	3.9	28,358	46.3
<i>is like</i>	27	26.2	15,940	26.0
<i>be like</i>	18	17.5	11,414	18.6
<i>are like</i>	13	12.6	5,101	8.3
<i>'re like</i>	4	3.9	4,279	7.0
<i>were like</i>	0	0	4,100	6.7
<i>'m like</i>	1	0.97	3,503	5.7
<i>been like</i>	0	0	2,437	4.0
<i>being like</i>	0	0	485	0.8
<i>am like</i>	0	0	179	0.3
	77	74.8	107,789	176.1

These data show that the *BE like* construction is more than twice as frequent in the COCA, with 176.1 occurrences per million compared to only 74.8 occurrences per million in the ELTCC. This difference further highlights the underrepresentation of *like*₁ in ELT coursebooks. We can also see discrepancies between the two corpora in the distribution of usage concerning the different conjugations of *BE*. The largest discrepancy is with the third person singular. In the COCA, the most frequent form is the contracted —'s *like*, occurring 52.3 times per million, while the standard *is like* is the most frequent form in the ELTCC, occurring 26.2 times per million. This suggests that coursebooks tend to underrepresent the usage of contractions. The next most frequent form of *BE like* in the COCA—the past-simple singular form *was like*—occurs 46.3 times per million, while the same form is considerably less common in the ELTCC, at only 3.9 occurrences per million. Only one form, *are like*, is more common in the ELTCC, where 12.6 occurrences per million compare to the COCA's 8.3 occurrences per million. Although the reason for this is unclear in the data, it seems that *BE like* comparisons in coursebooks tend to be more general (*are like*) than specific (*is like*). In addition, although these are among the least frequent in the COCA as well, four forms of *BE*—*were*, *been*, *being*, and *am*—are not represented at all in the ELTCC.

Another major difference between the ELTCC and the COCA with regard to the *BE like* construction is with the proportion of simile usage, which is much lower in the ELTCC. The ratios of simile usage in the COCA, reported in Chapter 5, Table 5.1, have been reproduced here and compared with the data from the ELTCC in Figure 9.2. The data in Figure 9.2 show that *BE like* simile usage in the ELTCC is notably lower than in any of the five sections of the COCA. As can be seen in the figure (and as it was

discussed in Chapter 5), simile usage is most common in fiction and least common in spoken English. A possible reason for the underrepresentation of *BE like* simile usage in the ELTCC may be the focus of this corpus on spoken language, with a smaller proportion of reading passages. This is typical in the majority of communicative ELT coursebooks.

Figure 9.2: Percentage of simile usage with the *BE like construction* in the ELTCC and the five COCA sections (spoken, fiction, magazine, news, and academic). Percentage of simile usage in ELTCC is calculated from total occurrences (77) of *BE like*. The data from the COCA are calculated from 500-concordance-line samples drawn from each of the five sections.



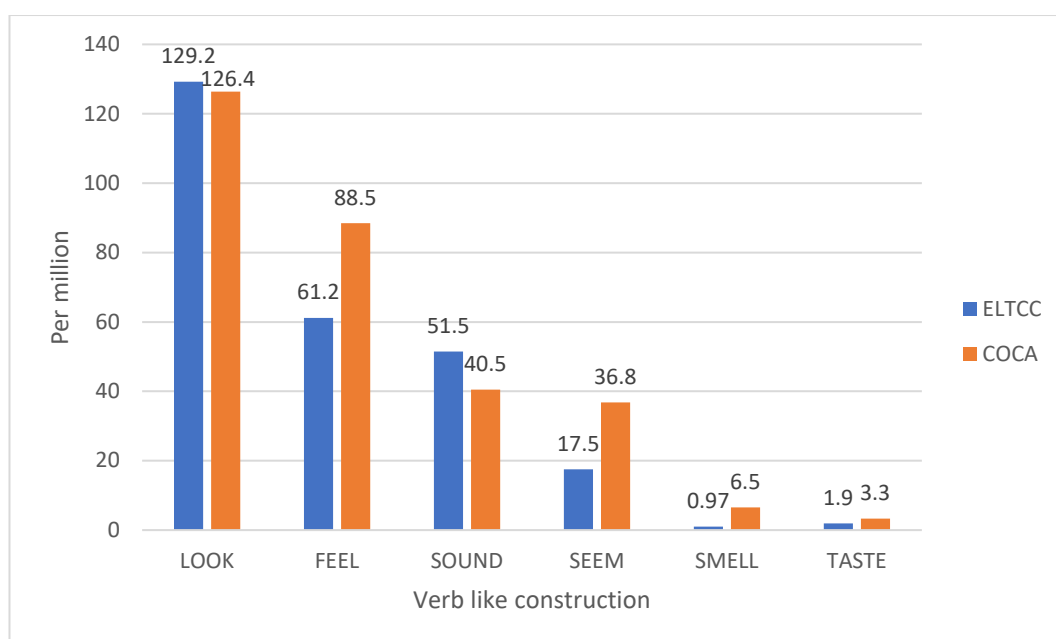
Most of the true *BE like* similes found in the ELTCC, including Examples 9.55–9.57, come from reading passages in the coursebooks.

- (9.55) Fish are **like** money in the bank. (CC[WorldEng1])
 (9.56) It's **like** finding a piece of gold in a river.
 (CC[AmEngFile3])
 (9.57) A life without dreams is **like** a garden with no
 flowers. (CC[FourCorners2])

Additionally, there are no overt explanations or pedagogic exercises on simile in the ELTCC—and there are no occurrences of the word *simile* in the ELTCC.

The PV like construction. In Chapter 6, it was shown that perception verbs (PV) are often used with *like* to form *PV like* phrases. It was also shown that there are two types of *PV like* constructions. One type features the polysemous *PV like* constructions *look like*, *feel like*, and *sound like*. The other type shows the mostly monosemous *PV like* constructions *seem like*, *smell like*, and *taste like*. Comparative analyses of the *PV like* constructions in the ELTCC and COCA show numerous discrepancies in frequency distributions for the different functions of the polysemous constructions, and overall frequency differences between the two corpora for the monosemous *PV like* constructions. This is shown in Figure 9.3.

Figure 9.3: Per-million occurrences of the six main verbs found in the *PV like* construction for the ELTCC and COCA



Look like is similar in both corpora, both in per-million frequency and as the most frequent *PV like* construction, although its usage is slightly higher in the ELTCC. *Feel like* is the second most frequent PV in both corpora but its occurrence is much higher per million in the COCA. *Sound like*

is the third most frequent in both corpora, but it has a higher frequency in the ELTCC. *Seem like* is much less frequent per million words in the COCA. Finally, *smell like* and *taste like* are very infrequent comparatively, but both are more frequent per million in the COCA. Table 9.4 illustrates.

Table 9.4: Frequency and per-million distributions of the six main verbs found in the *PV like* construction for the ELTCC and COCA

Verb like		ELTCC Frequency	(1,029,093) Per million	COCA Frequency	(612 M) Per million
LOOK	<i>look like</i>	100	97.2	27,562	45.0
	<i>looks like</i>	24	23.3	25,584	41.8
	<i>looked like</i>	7	6.8	21,072	34.4
	<i>looking like</i>	2	1.9	3,148	5.1
		133	129.2	77,366	126.4
FEEL	<i>feel like</i>	49	47.6	27,930	45.6
	<i>feels like</i>	5	4.8	7,170	11.7
	<i>felt like</i>	7	6.8	16,901	27.6
	<i>feeling like</i>	2	1.9	2,162	3.5
		63	61.2	54,163	88.5
SOUND	<i>sound like</i>	6	5.8	7,804	12.8
	<i>sounds like</i>	47	45.7	11,244	18.4
	<i>sounded like</i>	0	0	4,971	8.1
	<i>sounding like</i>	0	0	795	1.3
		53	51.5	24,814	40.5
SEEM	<i>seem like</i>	7	6.8	6,034	9.8
	<i>seems like</i>	10	9.7	8,989	14.7
	<i>seemed like</i>	1	0.97	7,455	12.2
	<i>seeming like</i>	0	0	64	0.1
		18	17.5	22,542	36.8
SMELL	<i>smell</i>	1	0.97	1,004	1.6
	<i>smells</i>	0	0	1,230	2.0
	<i>smelled</i>	0	0	1,508	2.5
	<i>smelling</i>	0	0	221	0.4
		1	0.97	3,963	6.5
TASTE	<i>taste</i>	1	0.97	666	1.1
	<i>tastes</i>	1	0.97	737	1.2
	<i>tasted</i>	0	0	568	0.9
	<i>tasting</i>	0	0	44	0.07
		2	1.9	2,015	3.3

In addition to the differences in overall frequencies between the two corpora, the frequency distributions of the different senses are notably different for all three polysemous *PV like* constructions.

Polysemous *PV like* constructions—LOOK like. Based on the 133 occurrences of *LOOK like* in the ELTCC, the vast majority (126 occurrences, or

94.7%) of *LOOK like* usage in this corpus is concerned with visual description. Half of these occurrences appear in the form of questions, with Example 9.58 occurring 12 times and Example 9.59 occurring nine times. The other half of the visually descriptive occurrences of *LOOK like* in the ELTCC consist of declarative sentences that mostly focus on describing what people look like (see Examples 9.60 and 9.61). The majority of these declarative statements are literal descriptions or comparisons, with only five sentences that could be classified as true similes (as with Examples 9.62 and 9.63).

- (9.58) What does he **look like**? (CC[FourCorners2])
- (9.59) What does it **look like**? (CC[EngFirstHandAcc])
- (9.60) She **looks like** a businesswoman.
(CC[CutEdgePreInt])
- (9.61) As you can see, I don't **look like** my brother.
(CC[SideBySide1])
- (9.62) 4. **Look like** a zombie b) look too pale
(CC[NiceTalkWithYou2])
- (9.63) Sarah says Matt's photos **look like** paintings.
(CC[Touchstone2])

Additionally, several occurrences of visually descriptive *LOOK like* are used for the pedagogic functions of providing instructions (often imperative sentences, as seen in Example 9.64) and explaining lexical items (as in Example 9.65).

- (9.64) Describe what it **looks like** and what it represents and explain why you like it.
(CC[WorldEng3])
- (9.65) You can use the verbs look and **look like** to talk about a person's appearance. (CC[AmEngFile3])

The remaining 5.2 percent of *look like* usage in the ELTCC is concerned with situational evaluation, as shown in Examples 9.66–9.68.

- (9.66) Oh, it **looks like** rain. (CC[FourCorners3])
- (9.67) So far, it's **looking like** a brilliant one.
(CC[Touchstone3])
- (9.68) I actually saw a movie about a flight attendant,

and it **looked like** something I could do.
(CC[AmEngFile4])

Comparing these findings to the corresponding COCA data (see 6.3), we can see a very large discrepancy in the frequency of usage of the two main senses. Visual sense description is overrepresented in the ELTCC (94.7% compared to 67% in the COCA) and, conversely, situational evaluation is underrepresented (5.3% compared to 33% in the COCA). While there are specific pedagogic explanations of visually descriptive *LOOK like* in the ELTCC (see Example 9.65 above), no similar pedagogic explanations of *LOOK like* for situational evaluation were identified in the data.

Polysemous PV like constructions—FEEL Like. There are 63 instances of *FEEL like* in the ELTCC; nearly half of these (30 occurrences, or 47.6%) are however associated with *like*_{2.3} (*in the mood for*), as discussed in section 9.4.1.3. Of the remaining 33 occurrences of *FEEL like* in the ELTCC, 15 (23.8%) are used for situational evaluation. See Examples 9.69–9.71.

(9.69) Sometimes, I **feel like** I'm not doing enough exercise and I spend too much time sitting at a desk. (CC[Passages2])

(9.70) It **feels like** a huge waste of time.
(CC[ComStrat3])

(9.71) It **feels like** she's judging me.
(CC[FourCorners3])

Next, there are 14 occurrences (22.2%) of *FEEL like* used to describe feelings, as seen in Examples 9.72–9.74.

(9.72) Oh, I always **feel like** this before I take a trip. (CC[WorldEng3])

(9.73) For a while, he **felt like** a kid again.
(CC[SideBySide3])

(9.74) I **feel like** I am loved. (CC[Passages1])

Only 4 occurrences (6.3%) of *FEEL like* in the ELTCC are used to describe tactile sensation, as in Examples 9.75–9.77.

- (9.75) It **feels like** a coin. (CC[AmEngFile4])
 (9.76) This material **feels like** silk. (CC[AmEngFile4])
 (9.77) It **felt like** my leg was still there.
 (CC[AmEngfile3])

Comparing these findings to the corresponding COCA data reveals further discrepancies in the frequency distribution of *PV like* usage between the two corpora. As mentioned previously, the usage of *FEEL like* to mean *in the mood for (like_{2.3})* is highly overrepresented in the ELTCC, accounting for 47.6 percent of all *FEEL like* usage. By contrast, this usage only accounts for 10.2 percent of *FEEL like* usage in the COCA. Due to this major discrepancy, both the *FEEL like* function of situational evaluation and that of describing feelings are underrepresented in the ELTCC. They account for 23.8 percent and 22.2 percent of usage respectively, compared to 50.2 percent and 33.2 percent respectively in the COCA. Only the least frequently used function of *FEEL like*, the description tactile sensation, is treated similarly in the two corpora, accounting for 6.3 percent of usage in the ELTCC and 6.4 percent of usage in the COCA.

Polysemous PV like constructions—SOUND like. There are 52 occurrences of *SOUND like* in the ELTCC, and the majority of these (80.7%, or 42 occurrences) are concerned with situational evaluation. There are 13 occurrences of Example 9.78.

- (9.78) That **sounds like** fun! (CC[EngFirstHand1])
 (9.79) That **sounds like** hard work. (CC[EngFirstHand1])
 (9.80) That **sounds like** a pretty big problem.
 (CC[WorldEng2])

The remaining 10 occurrences of *SOUND like* are used to describe aural sense, accounting for 19.2 percent of *SOUND like* usage. Three of these occurrences are used for pedagogic description, as shown in Example 9.83, which is focused on explaining reduced, casual forms of words for listening practice.

- (9.81) She **sounds like** Mariah Carey. (CC[Touchstone2])
 (9.82) It **sounds like** thunder to me. (CC[AmEngFile4])
 (9.83) Notice how they **sound like** hafta, hasta, and
 gotta. (CC[WorldEng2])

As with the other polysemous *PV like* constructions, the ELTCC and COCA have very different frequency distributions for the main functions of *SOUND like*. Situational evaluation is overrepresented in the ELTCC, accounting for 80.7 percent of *SOUND like* usage, compared to 53 percent in the COCA. Conversely, the description of aural sense is underrepresented, at 19.2 percent in the ELTCC but 47 percent in the COCA.

Monosemous PV like constructions—SEEM like. There are only 18 occurrences of *SEEM like* in the ELTCC, all of which consist of declarative statements, as shown in Examples 9.84–9.86.

- (9.84) Although the story may **seem like** it's for
 children, adults can really enjoy it too.
 (CC[Touchstone3])
 (9.85) It **seems like** a very selfish choice.
 (CC[Passages2])
 (9.86) It **seemed like** a good deal. (CC[Passages2])

The *SEEM like* construction is twice as frequent in the COCA, occurring 36.8 times per million words as compared with the ELTCC's 17.5 times per million.

Monosemous PV like constructions—TASTE like. There are only two examples of *TASTE like* in the ELTCC, as follows:

- (9.87) What does it **taste like**? (CC[Touchstone3])
 (9.88) This **tastes like** tea, not coffee.
 (CC[AmEngFile4])

Although *TASTE like* is a low-frequency construction in both corpora, again there is an underrepresentation in the ELTCC, where it occurs 1.9 times per million. It is slightly more frequent in the COCA, with 2,014 total occurrences, or 3.3 times per million words.

Monosemous PV like constructions—SMELL like. *SMELL like* is the least common *PV like* construction in the ELTCC, with only one example in the entire ELTCC, as follows:

(9.89) It doesn't really **smell like** anything.
(CC[AmEngFile4])

In the COCA, there are 3,957 instances of *SMELL like*, occurring 6.5 times per million words—nearly seven times the frequency of *SMELL like* in the ELTCC.

9.4.2.2: *Like*_{1.3} (*For Example*)

Four percent (20 occurrences) of the *like* data sample from the ELTCC is associated with the *like*_{1.3} sense of providing an example. In addition to declarative and interrogative sentences (see Example 9.90 and 9.91), this function of *like* is commonly used for pedagogic description with 33 occurrences of the pattern *VERB expressions like NOUN* (as seen in Example 9.92).

- (9.90) The students wanted healthier food **like** fruit
and yogurt. (CC[FourCorners3])
(9.91) **Like** what? (CC[WorldEng1])
(9.92) Use expressions **like** these to show surprise in
informal conversations: (CC[Touchstone2])

Instances of *like*_{1.3} are more common in the COCA, with 36 occurrences identified in the data sample, accounting for 7.2 percent of *like* usage. However, there were only 11 instances of *VERB expressions like NOUN* in the entire COCA, occurring only 0.02 times per million, compared to 32.1 times per million in the ELTCC. This fits with the common usage of language for pedagogic description in coursebooks.

9.4.2.3: *Like*_{1.2} (*Same As*)

In the COCA data, there is an equal proportion of *like*_{1.2} and *like*_{1.3}. In the ELTCC data, however, *like*_{1.2} (*same as*) is underrepresented, with only the following three occurrences:

- (9.93) **Like** credit cards, these tiles were imprinted with the owner's name and credit limit, and the name of the bank. (CC[AmEngFile4])
- (9.94) "**Like** many women, I am a size 16, and sometimes my vacation photos are not as flattering as I would like," said Sally Cranham, 24, a professional singer who tried out the camera. (CC[AmEngFile4])
- (9.95) I did my homework after dinner, **like** always. (CC[AmEngFileStart])

Usage of *like*_{1.2} is 12 times more frequent in the COCA, with 36 occurrences accounting for 7.2 percent of the data sample.

9.4.3: *Like*₃—Pragmatic Use

Pragmatic usage of *like* in the ELTCC is very rare. There are no instances of *like* being used as a discourse marker (*like*_{3.2}), and only one instance of *like* being used as a filler (*like*_{3.1}) was identified in the 500-concordance-line sample taken from the ELTCC. This instance (Example 9.96) accounted for only 0.2 percent of *like* usage. In addition, a search of the entire corpus identified four additional occurrences of *like*_{3.1}, all of which followed the word *mean*, with *like* separated by commas (see Examples 9.97 and 9.98).

- (9.96) I mean, it's cool you guys want to go and all, but the bands are **like** really old and kind of boring. (CC[EngFirstHand2])
- (9.97) I mean, **like**, what were you guys like? (CC[EngFirstHand2])
- (9.98) You mean, **like**, make it? (CC[EngFirstHandSucc])

Unsurprisingly, this function of *like* was found to be more frequent in the COCA, with 17 occurrences in the data, occurring for 3.4 percent of the

sample. A search of the entire COCA found 243 instances of *PRON mean, like*, and the majority occurred in the spoken section.

9.4.4 Premodifiers Used with *Like*

As previously discussed in 9.4.1.1, the most frequent premodifier used with *like* in ELT coursebooks is *really*, with 99 instances of *really like* in the ELTCC, or 96.2 times per million.

- (9.99) I **really like** Lady Gaga. (CC[AmEngFileStart])
 (9.100) I don't **really like** greasy food. (CC[NewInt1])
 (9.101) What is one extreme sport you would **really like** to do? (CC[ComStrat3])
 (9.102) It was **really like** an out-of-body experience, I guess. (CC[AmEngFile4])

The *really like* construction is much less frequent in the COCA, with 4,282 occurrences, or seven times per million words. While verbal usage of *really like* is the most frequent usage in both corpora, it is less frequent in the COCA, accounting for 77.6 percent of usage compared to 99 percent in the ELTCC. Conversely prepositional usage of *really like* is underrepresented in the ELTCC, where it accounts for only one percent of usage as compared with the COCA's 22 percent.

There are 16 instances of *just like* in the ELTCC, and the majority of these involve prepositional usage of *like*, with *just* taking on the meaning of *exactly* (see Examples 9.103 and 9.104). The remaining three occurrences of *just like* involve verbal *like*, with *just* being used to mean *only* (Example 9.105).

- (9.103) **Just like** us! (CC[FourCorners2])
 (9.104) The aliens looked **just like** creatures from a TV show. (CC[Passages1])
 (9.105) 66 percent of Japanese **just like** to relax. (CC[EngFirstHandSucc])

Instances of *just like* are much more frequent in the COCA, with prepositional *just like* occurring 33.3 times per million and verbal *just like* occurring 11.8 times per million. In contrast, prepositional *just like* and verbal *just like* occur in the ELTCC 12.6 and 2.9 times per million respectively.

Additionally, there are no instances of either the phrase *just like that* or the pattern *just like NOUN to VERB* in the ELTCC. As discussed in Section 7.4.2, these occur rather frequently in the COCA.

Instances of the comparative *more like* are infrequent in the ELTCC, with only seven instances (represented in Examples 9.106–9.108). No instances of either the phrase *that's more like it* or the phrase *more or less like* were found in the data. Nor was the pattern *BE more like it* found. These were discussed in 8.4 and 8.5.

(9.106) Is your life **more like** Sasha's or Sam's?

(CC[EngFirstHand1])

(9.107) Do you look **more like** your mother or father?

(CC[EngFirstHand1])

(9.108) To the untrained ear, the language sounds **more like** humming than speech. (CC[AmEngFile4])

In the COCA, the comparative *more like* construction is much more frequent with 13,667 occurrences, or 22.3 times per million, compared to only 6.8 times per million in the ELTCC.

The superlative *most like* is more frequent in the ELTCC, with 28 instances; 20 of these (71.4%) are verbal, as seen in Examples 9.109 and 9.110. The other eight (28.6%) are prepositional, as in Examples 9.111 and 9.112.

(9.109) Which city would you **most like** to visit? X6

(CC[AmEngFile4])

(9.110) Which person would you **most like** to meet?

(CC[Touchstone3])

(9.111) Which person dresses the **most like** you?

(CC[Passages2])

(9.112) In your family, whose personality is the **most like** yours? (CC[EngFirstHand1])

The *most like* construction is notably less frequent in the COCA, with 520 instances or a rate of only 0.85 times per million, compared to 27.2 per million in the ELTCC. The ratio of prepositional *like* and verbal *like* in the *most like* construction is more evenly divided in the COCA, with 56 percent and 41 percent respectively. Additionally, there are 14 occurrences in the COCA of *most like* being used to mean *most likely*, and this usage is not represented in the ELTCC.

There are five instances of *least like* in the ELTCC, with four involving verbal *like* (as with Example 9.113) and one involving prepositional *like* (Example 9.114). There are no instances of *less like*.

(9.113) What do you **least like** about flying?

(CC[AmEngFile4])

(9.114) Who is **least like** you? (CC[Passages2])

The *least like* construction is less frequent in the COCA, with 115 occurrences, a rate of only 0.19 times per million words, as compared with 4.6 times per million words in the ELTCC. Despite this, however, there is a greater proportion of prepositional *least like* usage in the COCA, with 54 percent prepositional and 41.7 percent verbal usage, compared to the ELTCC's 20 percent and 80 percent prepositional and verbal usage respectively. Additionally, there are five occurrences in which *least like* is used as a filler phrase (*like*_{3.1}) in the COCA. Meanwhile, this usage is unsurprisingly absent from the ELTCC, as there are very few instances of pragmatic usage of *like* in ELT coursebooks.

9.5: Pedagogic Implications and Applications

9.5.1: Implications

While much has been said in the literature about the pedagogic inadequacies of the more traditional grammatical and lexical approaches to language, I have presented further evidence to support these arguments in Chapters 7 and 8. Using just one word as an example, it was easy to see how relying on a reference grammar can be insufficient and at times even misleading, in that several grammatical constructions related to the premodification of *like* were not included. Furthermore, neither the reference grammar nor the dictionary was helpful in determining the meaning of several frequent patterns and phrases found to occur with *like*. These observations add support to the argument that grammars and dictionaries need to be revised (and perhaps even reimagined) with the aid of corpus-based and corpus-driven techniques and methodology. However, this research shows just how complex language is, and how difficult it is to explain every detail of the lexicogrammatical system. This is even more apparent considering that the definition for the phrase *(and) just like that* could not even be found in the *Collins COBUILD Learner's Dictionary* (2003; *CCLD*). Although this is of course a corpus-based dictionary, it offered no insight on the phrase, under either of the headings for *just* and *like*.

The corpus evidence presented in Section 9.4 clearly shows major discrepancies between the treatment of *like* in ELT coursebooks and how it is used in the natural English reflected in the COCA. The most frequent usage of *like* in the COCA, *like*₁ (*similarity*), is notably underrepresented in the ELTCC—which, conversely, overrepresents the usage of *like*₂ (*enjoy/want*). In addition, the frequency distributions of sense usage for the polysemous *PV like* constructions vary widely between the ELTCC and the

COCA. Coursebook language featuring *like* was also found to involve the use of questions and pedagogic description much more often than in natural language use, and this is assumed to be true for coursebook language in general.

The underrepresentation of a highly frequent lexical item can have pedagogical consequences for learners. Such consequences include those I have observed in my own teaching context. Noticing that some of my first-year university students misunderstood the *what BE N like* construction, I carried out an informal research project in the classroom. I wanted to investigate whether this problem affected only a few students, or a larger proportion of the class. In the study, three groups of students were instructed to write a short answer to the question *What is your hometown like?* The question was dictated, and the students wrote their answers on pieces of paper which were collected for analysis.

Table 9.5: Results of informal classroom research, with frequency of answer types, from first-year university students' answers to the question *What is your hometown like?*

	Describe Hometown	Like About Hometown	Hometown Name
Class 1 (Engineering I)	12	17	5
Class 2 (Engineering II)	16	19	1
Class 3 (Medical / Agriculture I)	13	21	0

The majority of responses seemed to incorrectly answer the question. Examples 9.115–9.117 below are representative examples of the students' answers.

(9.115) I **like** beautiful nature in my hometown.

(9.116) I **like** my hometown because my hometown is safety and have good nature.

(9.117) I **like** the beautiful river.

These errors are clearly based on a lack of awareness of the various functions of like, as the students are misunderstanding prepositional *like* as if it were verbal *like*. This may be due in part to the overrepresentation of verbal *like*₂ in coursebooks. Somewhat surprisingly, this overrepresentation of verbal *like* in coursebooks is also found in *Touchstone*, a corpus-based ELT coursebook series included in the ELTCC. Despite its advertising, by which the series is meant to be understood as corpus-based and therefore made up of more natural language use, *Touchstone* did not appear suitable for any correction of my students' misunderstanding. An analysis of the three *Touchstone* coursebooks revealed an even stronger overrepresentation of verbal like than the ELTCC average, with 91.5 percent *like*₂ usage and only 8.5 percent *like*₁ usage.

9.5.2: Applications

A solution to the problem of under- and overrepresentation of lexical items in teaching materials can easily be achieved by consulting appropriately representative corpora and incorporating the principles set out by Sinclair and Renouf's (1998) and their notion of the lexical syllabus. In the case of our target word *like*, coursebook authors could focus on the most frequent patterns of usage, and the most frequent collocations—of the most frequent lexical items. This way, students would have much more exposure to *like*₁ (*similarity*) usage, and would therefore be less likely to misunderstand a question like the one I used: *What is your hometown like?*

As noted in Peppard (2016 27–28), one criticism of the lexical syllabus is that an overreliance on corpus data and frequencies can lead to coursebooks that are not engaging. Such coursebooks might instead be marked by an appearance that Marc Helgesen, a coursebook author, called “machine-assembled” (personal communication, 2015). Helgesen argued

that coursebooks, and oral communication coursebooks in particular, are designed to do a lot more than present vocabulary, grammar, and functions. They are also designed to organize tasks and interactions that are essential in creating a friendly classroom atmosphere. He noted that most books are meant to break the ice and get students talking to each other early in the course; discussing likes and dislikes is a good way to do this. This probably contributes to the overrepresentation of the verbal usage of like in the ELTCC and the coursebooks. In his discussion of graded readers, Viney (2015) made a similar argument against the exclusive use of word lists based on frequency for vocabulary selection:

Teachers have said, “It’s easy. Why don’t they just use frequency counts?” But it’s not easy at all. The demands of graded readers are different. You need words like said, told, whispered, shouted, screamed as soon as the past tense is available. Words for dramatic events—gun, fire, sword, murder—might be more important than salt and pepper. (para. 8)

Sinclair and Renouf (1988) addressed this line of criticism regarding the lexical syllabus. While advocating the importance of frequency for vocabulary selection, they acknowledged that frequency should not be the only indicator used, noting that the 200 most frequent words are function words, semantically empty. They conceded that words outside the target frequency band would need to be added, such as those relating to everyday domestic reality and classroom procedure words. Another important factor to consider here is with natural sets of vocabulary items which should be presented together despite some of the items having lower frequencies. Stubbs (2001, 20) clearly exemplifies this point with the days of the week, in which the words *Saturday* and *Sunday* occur in a 150-million-word corpus more than twice as often as *Tuesday* and *Thursday*. It seems, then, that there is a place, however limited, for subjectivity as well as lower-frequency

items in the selection process constructing applications of the lexical syllabus, but I maintain that there must be a specific need for any item that is chosen subjectively, such as the inclusion of low-frequency words as part of a larger, higher-frequency set. In Koprowski's (2005) critique of vocabulary selection, for example, he lists *do judo* as an item that is not particularly useful to learners. However, it may be of use in Japan, where *I played judo [or, e.g., karate] in high school* features a common error made by university students talking about their past extracurricular activities.

A final point to consider regarding the usefulness of frequency as the main criterion for the selection of vocabulary items in EFL materials and coursebooks concerns the effects of genre/register variation on frequency. Gablasova et al. (2017, 167) note that there is ample evidence showing that genre/register, as reflected in corpora and subcorpora, is a major predictor of language variation, and this has been demonstrated across various aspects of language use (e.g. morpho-syntactic, pragmatic, lexical, and formulaicity). In the current study, I have shown that certain *like* constructions occur with notably different frequencies across different genres and this raises two important questions: Does divergence from the overall mean frequency of *like* from the COCA always indicate over/underuse when compared to the ELTCC, and should these discrepancies simply be considered as appropriate for the EFL coursebooks register? These questions highlight the importance of selecting appropriate corpora or subcorpora that best represent the targeted genre/register.

While the above-mentioned criticisms of the lexical syllabus should certainly be taken into consideration, I would argue that these points are moot with regard to *like*. This study has shown that *like* plays an integral role in natural English use, across several genres, and while it has three

main functions (*like*₁ [*similarity*], *like*₂ [*enjoy/want*], and *like*₃ [*pragmatic*]), *like*₁ (*similarity*) is by far the most frequently used general function across all genres. However, this use is significantly underrepresented in ELT coursebooks, and I have shown that this likely contributes to the high level of student error associated with *like*.

In order to incorporate a lexical-syllabus component into a coursebook, the coursebook writers would first need to check their overall vocabulary representation. This could be done retrospectively, by running the text through concordance software such as AntConc (see Chapter 4) to produce a wordlist that could be checked alongside a corpus-based frequency list like the one available on the COCA website. If it were discovered, for example, that the prepositional (*like*₁) use of *like* is underrepresented in the coursebook—as it often is—the authors would need to take measures to remedy this underrepresentation. This could be done in one of three ways. First, the authors could revise the coursebook and simply add components that incorporate *like*₁ use. Second, they could add a language point box, or some similar element, alongside the verbal (*like*₂) use of *like*. This would have the effect of drawing students' attention to the more frequent *like*₁ usage, with a warning concerning the common error of confusing the two forms of *like*. Finally, the authors could include exercises to help the students discern the different functions of *like*, such as a guided DDL exercise similar to the one shown in Subsection 2.6.2.

9.6: The Lexical Syllabus As Solution

This chapter has provided an overview of the current state and main trends regarding the treatment of phraseology in ELT coursebook design. I compared and discussed frequency distributions for the main senses and functions of *LIKE* in the COCA and the ELTCC, comparing the main COCA

findings from Chapters 5 through 8 with the corresponding data from the ELTCC. Finally, I discussed the pedagogic implications of these findings and made some recommendations on how coursebook designers might be able to incorporate corpus-based pedagogic applications of the lexical syllabus and data-driven learning into future coursebooks. This would allow them to better deal with vocabulary in general, and with high-frequency words such as *LIKE* in particular.

I have made the argument in this chapter that the vocabulary selection process for ELT coursebooks is usually flawed. I have provided examples to back up Koprowski's (2005) claim that the selection process is often conducted subjectively and without reference to corpus data. Although Helgesen (personal communication, 2015) and Viney (2015) make valid points in defence of ELT materials, it seems that there is certainly room for improvement. Coursebook writers can address the selection problem by following Sinclair and Renouf's (1988) recommendation of presenting students with the most common patterns and collocations of the most frequent words in the language. While this should be the main factor in determining the lexical syllabus for the majority of the items in a coursebook, there will still be a need for a certain number of subjectively chosen items. However, careful consideration should be made regarding each item's usefulness and, in the case of such a high-frequency, error-prone word as *like*, coursebook writers would be well advised to ensure that they are representing the most frequent patterns of usage. As for presenting students with the lexical items found in the lexical syllabus, I have argued that the integration of DDL exercises could prove to be beneficial (see also Peppard 2014). These integrated DDL exercises, which have been shown to be effective as consciousness-raising activities (e.g. Johns 1991, 1994; Tian 2004; Bolton 2007a,

2007b, 2008; Peppard 2014; Cobb and Boulton 2015), can be especially useful for helping students make sense of high-frequency words, such as *like*, that have numerous functions and patterns of usage.

10

Conclusion

Like, Corpus Findings, and ELT Coursebooks

10.1: *Like*, Corpora, and ELT Coursebooks

In this thesis, I have conducted a corpus-based study of the highly frequent word *like* from a lexical syllabus perspective. This involved three main steps. First, I examined data from the Corpus of Contemporary American English (COCA) to determine the most frequent functions of *like*, including its most frequent patterns, phrases, and collocates, to understand how it is used in everyday English. Next, I compiled the English Language Teaching Coursebook Corpus (ELTCC), a specialized pedagogic corpus of ELT coursebooks, and examined the corresponding data to determine how *like* is treated in commercially available teaching materials. Finally, I compared the findings from these two corpora to determine how closely—if at all—ELT coursebook writers adhere to the principles of the lexical syllabus. This is important because the lexical syllabus states that a language syllabus should focus on the most frequent words of the language and include their most frequent functions and patterns of usage.

This concluding chapter is divided into three main sections. In Section 10.2, I review the four main results of the study based on a review of all the findings from Chapters 5 through 9. In Section 10.3, I discuss both the

strengths and the limitation of this study for my thesis; and finally, in Section 10.4, I make some suggestions for further research.

10.2: Summary of Findings

In reviewing the data and summarizing the findings from Chapters 5 through 9 of this corpus-based study of *like*, four main results can be observed:

1. Although it is often ignored in the literature, simile is very frequent, highly complex, and worthy of further study.
2. Sense verbs do not always behave as true sense verbs when paired with *like*, and they are often used for situational evaluation.
3. Traditional lexical and grammatical approaches are often inadequate and sometimes misleading.
4. *Like* is highly misrepresented in ELT coursebooks, which can lead to student error.

These four results will be discussed in more detail below with examples from the data presented in previous chapters. For convenience, their numbers are repeated here as originally presented.

10.2.1: Simile is Worthy of Further Study

In Chapter 5, initial analysis of the corpus data showed the *BE like* construction to be the most frequent of recurring *like* constructions in the COCA. Further analysis of this data revealed that the majority of *BE like* occurrences were used to form similes. Interestingly, however, in Chapter 3, it was clear that simile may have been examined in some detail in the literature, but that it has rarely been the main focus of study.

As it turns out, descriptions of simile are most commonly included in discussions of metaphor, discussions in which the two phenomena are compared, with simile often dismissed as a subset, or lower form, of metaphor (Hanks 2005, 1; Veale and Hao 2007, 683). However, this incongruence

seems misguided, and it seems so because there are claims that simile occurs just as often as metaphor (Fadaee 2011, 22). Other evidence suggests it may even occur more often (Hanks, 1). Meanwhile, categorization and career of metaphor (two of the main models of metaphor) both treat simile as a separate and independent phenomenon.

Furthermore, with the notable exceptions of Hanks, Moon (2011b), and Wikberg (2008), there are relatively few corpus-based phraseological studies of simile in the literature. The examples provided in many studies are simplified and removed from their context (Wikberg, 129) and there have been even fewer corpus-based studies of simile that examine genre differences. By providing an in-depth, cross-genre, and corpus-based analysis of simile usage in the current study, I feel that I have contributed to the perceived gap in this field where simile-specific research is lacking. The results from Chapter 5 show that simile is very frequent, highly complex, and indeed a worthy subject of investigation in and of itself, without the need for a comparison to metaphor.

While Fadaee (2011, 22) noted that similes are likely found in most genres, if not all—including prose, poetry, and conversation—most of the previous studies showed that similes are most commonly found in fiction (Wikberg, 129). The study presented in this thesis supports those claims and adds further evidence for the notion that simile usage is a ubiquitous function of everyday language use. The evidence, while confirming that simile use is most frequently deployed in fiction, shows that *BE like* simile usage is highly frequent across all five genres of the COCA (see Section 5.2). The data for this study also showed that similes are least common in spoken English. However, the most frequent use of lexicalized similes occurs in

spoken English, where they are likely used to facilitate speed in real-time language processing.

Another factor that has highlighted the complexity of simile usage concerns the interactions between simile type and genre. Analysis of the corpus data revealed three notable trends. As I showed in Chapter 5 (Example 5.7), nominal predicative similes are more frequent in academic writing. Clausal similes as shown in Example 5.13 are less frequent in academic writing, and verbal predicative similes (as in Example 5.10) are more frequent in magazines.

- (5.7) Globalization **is like** an economic freight train.
If one is prepared to jump on board, then one
can go far—if you pay the price for the ticket.
(ACAD[ArabStudies])
- (5.13) It **was like** he stuck a knife into my heart.
(SPOK[ABC_20/20])
- (5.10) Treating them like puzzles **is like** trying to
solve the unsolvable—an impossible challenge.
(Magazine)

These findings suggest that genre affects not only the frequency of simile usage but also the type of simile. Based on the data and my qualitative corpus-based speculation, it seems likely that the academic preference for using objects over actions or situations for vehicles of similes is based on the need to describe things in more concrete terms. Academic writers need to avoid obscurity and ambiguity. They also tend to follow conventions of formality in their prose. The avoidance of clausal similes in academic English is most likely due to the use of clausal *like* similes being frowned upon in prescriptive grammars, as they are considered unrefined and colloquial. Finally, the tendency for magazines to favor verbal predicative similes is possibly the result of magazine writers aiming for a more dynamic,

experiential effect with their descriptions. In this way, they may make their writing as interesting as possible for the reader.

With regard to the main functions and rationale behind simile usage, previous research has centered around explanation and description, including the expression of emotion, and for making descriptions more vivid and entertaining. The study for this thesis provides further insight into the function and rationale behind the usage of *BE like* similes, showing that they are often used for the purposes of elaboration. This is evident in the finding that, in a large number of similes, the usual slot where the target should be located consists of the pronoun *it* (see Example 5.55). As I observed with this example in Chapter 5, the actual target of the simile is located in the preceding text.

(5.55) Because there are so many fine lines it's possible to detect even minute shifts. "It's **like** holding the star up to a piece of graph paper," McCarthy says. (MAG[Smithsonian])

In these cases, the speaker has already spoken about the target and is following up, or elaborating, with the use of a descriptive simile.

In Chapter 5 I also observed that while previous studies have stated that simile is commonly used for evaluation (Wikberg, 137), the data in the COCA suggest another common use. In Example 5.53 I provided evidence showing that *BE like* similes are often used for the more specific function of stating opinions.

(5.53) Bankers **are like** trapeze artists. The greater the net you put under them, the greater the risk they'll be willing to take. (SPOK[ABC_20/20])

10.2.2: Sense Verbs Are Not Always Sense Verbs

In Chapter 6, I looked at the *VERB like* construction, which I had found to be the second most frequent recurring *like* construction in the COCA. I found that the vast majority of verb usages in this construction consist of the perception verbs (PV): *LOOK*, *FEEL*, *SOUND*, *SEEM*, *SMELL*, and *TASTE*.

The corpus data showed that the six verbs used in this *PV like* construction form two distinct groups. The monosemous group consists of the three least frequent PVs—*SEEM like*, *SMELL like*, and *TASTE like*. Meanwhile, the polysemous group consists of *LOOK like*, *FEEL like*, and *SOUND like*, which account for the vast majority of all *PV like* usage in the COCA. While *LOOK*, *FEEL*, *SOUND*, *SMELL*, and *TASTE* are generally referred to as the five sense verbs, because they are used for describing sensory experience, I found their usage more complicated. They are paired with *like* for a large portion of their usage, and it turned out that most of these sense verbs do not behave as true sense verbs. Instead, the most frequent polysemous sense verbs—*LOOK*, *FEEL*, and *SOUND*—take on both literal and figurative applications. All three of the constructions in which these verbs are paired with *like* share the related function of situational evaluation, showing nuanced differences depending on the verb. In this regard, *LOOK*, *FEEL*, and *SOUND* are much more similar to the functioning of *SEEM*, as opposed to the true sense verbs *SMELL* and *TASTE*. Here are some corpus examples from Chapter 6.

- (6.30) But, right now, it **looks like** we got pretty lucky. (NEWS[Houston])
- (6.44) We **feel like** consumer confidence is rising. (NEWS[USAToday])
- (6.75) It's **sounding like** it's going to be a game time decision. (SPOK[Fox_Susteren])
- (6.6) Everything **seems like** a dream. (SPOK[NPR_ATC])

In the cases of *FEEL like* and *SOUND like*, this figurative usage occurs even more frequently than the literal description of sense. Additionally, *FEEL like* was found to be the most polysemous, with the additional functions of describing one's feelings and describing mood or wants using the *like*_{2,3}, *FEEL like* construction.

(6.53) We **feel like** kids in a candy shop.
(NEWS[Atlanta])

(6.58) But I don't **feel like** doing that project, to
tell the truth. (FIC[KansasQ])

10.2.3: Traditional Approaches Are Often Inadequate

Proponents of corpus-based and corpus-driven approaches to language description have long argued that, when dealing with natural language use, the more traditional and prescriptive grammatical and lexical approaches to language are often inadequate or misleading. In Chapters 7 and 8, I provided further evidence supporting these arguments by examining the most frequent premodifiers found to occur with *like* in the COCA. I compared these results with the language prescriptions found in a traditional dictionary (*Merriam-Webster*, accessed September 26, 2020) and a reference grammar (Downing and Locke 2006; D&L).

10.2.3.1: *Just Like*

In chapter 7, to demonstrate the problems that can arise when attempting to determine meaning at the single-word level, I looked at the intensifying premodifiers used with *like* in the COCA. The corpus analysis allowed me to identify *just* as the most frequent intensifying premodifier to be used with *like*. Compared to other premodifiers, it was most frequent by a large margin and, for this reason, I carried out a more detailed analysis of the *just like* construction. In this analysis I found that in the majority of *just*

like usage, *just* does in fact act as an intensifier. It shifts the meaning of *like* from *similar* (*like*_{1.1}) to *same* (*like*_{1.2}). Additionally, numerous patterns and phrases using *just like* were identified in the data with varying levels of lexicalization.

While many of the lexicalized *just like* patterns and phrases took on specialized meanings, they could still be classified within the grammatical framework outlined by D&L. They retained their compositional dictionary definitions. The following examples are all drawn from Chapter 7.

- (7.13) But now he says that Sean is **just like any other** kid on the block. (SPOK[NBC_Dateline]).
- (7.29) Her wish, **just like always**, was his command. (FIC[Bk:MistressNoMore])
- (7.34) Many are **people just like you**, chasing the American dream. (SPOK[NBC_Dateline])

However, two items in particular were identified in the data that were found to not fit within the traditional grammatical framework for intensifying pre-modifiers. First, the pattern *BE just like N* is a lexicalized construction that is used to describe someone's behaviour as being typical. This was seen in Example 7.53, which I will show here for review.

- (7.53) It was **just like him** to downplay his success. (FIC[Cosmopolitay])

This pattern has several variations, including *it BE just like NOUN*, *isn't it/that just like NOUN*. And it is often followed by a *to infinitive* complement. In all of these variations, there are no corresponding instances in the data that do not include *just*. So, in this case, *just* does not appear to act as an intensifier.

Next, the frequently occurring phrase (*and*) *just like that*, and its variation *VERB just like that*, appeared to be fully lexicalized phrases that would

not fit within the traditional grammatical framework of intensifying premodifiers. See Examples 7.74 and 7.479.

(7.74) **And just like** that he was done. (MAG[Essence])

(7.79) Mommy got sick and it **happened just like that**
and there was nothing anybody would do.
(FIC[Mov:AmericanPresident])

In this phrase, which is used to mean *quickly*, *suddenly*, or *easily*, *just* does not behave as an intensifying premodifier. While there are instances of *like that* in the data, this phrase takes on a different meaning, one which is demonstrative and deictic. Additionally, *just like that* was found in neither the traditional dictionary nor the more modern corpus-driven *Collins COBUILD Learner's Dictionary* (2003). Entries for both *just* and *like* missed this sense.

The *just like* construction also turned out to be used in its pragmatic *like*₃ form. Here too, *just* does not conform to the traditional grammatical rules of intensification. It is used as a filler (*like*_{3.1}), as shown in Example 7.79, where it is part of the extended filler phrase, *just like, you know*, and it is also used as a discourse marker (*like*_{3.2}), as shown in Example 7.81.

(7.86) What if they **just like** ignore me while I'm
saying it? (ACAD[Adolescence])

(7.88) And he was **just like**, "Tiffany, look where that
city is." (SPOK[ABC:20/20])

10.2.3.2: *Like* with Comparative and Superlative Premodifiers

In Chapter 8 I began by examining the functions of *like* with comparative and superlative premodifiers, focusing primarily on the comparative *more*. I had found *more*, after *just*, to be the second most frequent premodifier to occur with *like*. Grammatically based comparative/superlative constructions used with *like that* which are not included in D&L's grammatical framework include *as much*, *more or less* (used to mean *slightly similar to*), *at least*,

and *about as much*. In addition, while D&L (538) do include *in the least* with their list of attested grading modifiers known to work with prepositions, they seem to have missed an important feature of this construction. All ten occurrences of *in the least like*, as found in the COCA, are preceded by a negative, forming the phrase *not in the least like*. For illustration, I will present Example 8.20 once more.

(8.20) Why, it's **not in the least like** a ship.
(FIC[Analog])

While this usage is grammatically based on the comparative and superlative rules outlined by D&L, the phrase *not in the least like* seems to take on the meaning of *not at all like*.

The comparative *more* was found to be the most frequently used of all the comparative and superlative premodifiers used with *like*. One of the reasons for this high frequency has to do with the *more like* construction being used in two specialized constructions. While *more like it* is sometimes found in standard comparative statements, the most frequent usage of this construction is found in the *NOUN BE more like it* pattern. This was illustrated in Example 8.34.

(8.34) Dump is too clean a word, dive is too nice,
abomination **is more like it**.
(FIC[Mov: FreddysDead])

This usage consists of a formula that functions to state an improved or refined description of something stated previously. It is often presented as a stated opinion. Additionally, this usage of *BE more like it* often consists of creative word play by the speaker or writer, in which the improved description is made to sound very similar in form to the entity it is being compared to. This word play is often generated from, or shows some response to, an interlocutor.

Next, the fully lexicalized phrase *that/this BE more like it* is not included in D&L's grammatical framework for comparatives and superlatives. See Example 8.54.

(8.54) Now the stove lids were glowing. "**That's more like it**, eh? Uncle Labe said, shedding his coat.
(MAG[BoysLife])

While on the surface it also appears to be a straightforward comparative usage, closer inspection of the corpus data show that the pronoun *it* in these cases usually indicates no inferable item of comparison. According to D&L (487), for the comparison to be understood, an inferable item is expected. This construction is used more often as a statement of contentment than a comparison.

By comparing the corpus data with the traditional grammatical account of how intensifying premodifiers and comparative and superlative premodifiers work, I observed two main findings. First, considering the vast complexity of language, even the most comprehensive traditional grammars such as D&L's cannot cover the full spectrum of all grammatical items. This is evident in the number of grammatically based items found in the corpus data but not included in the comparative/superlative grammatical framework. Second, traditional grammars often fail to take phraseology into consideration and, because of this, frequently used lexical patterns and phrases may not be accounted for. This can be particularly detrimental to language learners.

10.2.4: *Like* in ELT Coursebooks

Finally, this thesis has shown that *like* is highly misrepresented in ELT coursebooks, as reflected in the ELTCC. From a pedagogical perspective,

the most significant result of this study involves the two major discrepancies discovered between the COCA and the ELTCC with regard to the treatment of *like*. These were outlined in Chapter 9. First, *like* was found to be nearly twice as frequent in the ELTCC compared to the COCA. Second, the frequency distributions of the various functions of *like* were found to be significantly different between the two corpora, with the two most frequent uses of *like* showing nearly opposite distributions. Usage of *like*₁ (*similarity*) is highly underrepresented in ELT coursebooks, with only 26.2 percent of *like* occurrences being used to show similarity. Meanwhile this is the most frequent usage of *like* in the COCA, where it accounts for 77.8 percent of total usage, and this underrepresentation was found to occur across all three sub-senses of *like*₁.

In contrast, with a nearly opposite frequency distribution, the data showed that *like*₂ (*enjoyment*) is highly overrepresented in coursebooks, accounting for 73.6 percent of usage in the ELTCC. This usage accounts for only 20.4 percent of *like* occurrences in the COCA. Looking at the sub-senses of *like*₂, I noted that the largest discrepancy between the two corpora involves *like*_{2.1}, meaning *fond of*, with 54 percent of *like* usage in the ELTCC showing this sense. By contrast, this usage accounts for only 15 percent of the COCA data. Usage of *like*_{2.2}, meaning *want*, is also overrepresented in the ELTCC, accounting for 19.4 percent of *like* usage—compared to just five percent in the COCA, this too was significant. Finally, while occurrences of *like*_{2.3} are relatively uncommon in both corpora, this usage was found to occur slightly more in the COCA sample, accounting for 0.4 percent of total *like* usage (compared to only 0.2 percent in the ELTCC sample).

Due to these significant misrepresentations of *like*₁ in the ELTCC, it was not surprising to find that the COCA's four most frequent patterns of

usage associated with *like* (as examined in Chapters 5 through 8) are underrepresented in the ELTCC. First, the ELTCC reflected that simile usage with the *BE like* construction was very infrequent in ELT coursebooks. The word *simile* did not occur at all, and there were no explicit descriptions or explanations of simile in the entire corpus.

Next, numerous discrepancies were found between the COCA and the ELTCC concerning the frequency distributions of the different functions for the polysemous *PV like* constructions. With *LOOK like*, visual sense description is overrepresented in the ELTCC, and conversely, situational evaluation is underrepresented. The usage of *FEEL like* to mean *in the mood for* (*like*_{2.3}) is highly overrepresented in the ELTCC, while both the situational evaluation function and that for the description of feelings with *FEEL like* are underrepresented. Only the least frequently used function of *FEEL like*, the description of tactile sensation, is treated similarly in the two corpora. With *SOUND like*, situational evaluation is overrepresented in the ELTCC, and the description of aural sense is underrepresented.

Regarding the premodification of *like*, there are very few instances of *just like* in the ELTCC, and there are no instances of any of the frequent patterns and phrases found in the COCA with *just like*. Conversely, the most frequent premodifier used with *like* in ELT coursebooks was found to be *really*, whereas the *really like* construction is much less frequent in the COCA. While verbal usage of *really like* is the most frequent usage in both corpora, it is less frequent in the COCA, and prepositional usage of *really like* is underrepresented in the ELTCC. Finally, instances of the comparative *more like* are infrequent in the ELTCC, and there are no instances of the phrases *that's more like it* or *more or less like*, nor any instances of the pattern *BE more like it*.

Taken together, these findings highlight a notable mismatch between corpus research and ELT coursebook design with regard to the treatment of the highly frequent word *like*, and this misrepresentation may be responsible for frequent misunderstanding and errors made by students. It is clear that, generally speaking, many ELT coursebook designers do not follow the principles of the lexical syllabus—at least not in the case of *like*. If these principles were followed, namely, a focus on the most frequent words and their most frequent patterns of usage, the frequency distribution of *like*₁ and *like*₂ would be reversed in coursebooks. Then, students would be more likely to understand the meaning of questions such as *What is your hometown like?*

10.3: Strengths and Limitations of This Thesis

10.3.1: Strengths

The study presented in this thesis has two main strengths.

First, because the research involves both quantitative and qualitative methods, it allows for a deeper and more detailed analysis than if it had been exclusively quantitative or exclusively qualitative. Starting with a more corpus-driven approach, I was able to identify the most frequent patterns of usage associated with *like*. Then, shifting to a more corpus-based approach, I was able to apply further analyses to construct a detailed understanding of how *like* behaves in these most frequent patterns of usage.

This hybrid approach has allowed for a depth of analysis that was well suited for addressing the research questions, which had led to numerous noteworthy findings. Furthermore, some of the main findings have corroborated previous research, while other findings may prompt a re-analysis of some previous studies, while still other findings have contributed to filling perceived gaps in the literature. For example, my findings in Chapter 6 have

supported the work of Viberg (1983), Sweetser (1990), and Winter (2019) with regard to the frequency of sensory description, confirming that visual sense description is most frequent, while smell and taste are least frequent. My research adds to the field, however, in that it has specifically examined the behaviour of these perception verbs (PV) when paired with *like*. This led to the finding that when paired with *like*, PVs do not always behave as true PVs but take on other functions. As for contradictory findings in this chapter, it was noteworthy that my findings concerning the lexicalized function of *FEEL like* to mean *in the mood for* contradict the findings of Hanks (2005), and for this reason, further analysis of the construction is merited.

Second, while I have gained (and presented) insight into the main functions of *like*, including additions to our understanding of simile usage and perception verbs, the findings of this study are directly applicable to the field of English language teaching. These findings can contribute to the improvement of design processes and needs assessments for teaching materials and lesson plans. I have shown that there is a very large discrepancy in how *like* is treated in ELT coursebooks and how *like* behaves in natural English use. My research also suggests that this discrepancy may be responsible for a high level of student error, in both production and reception of *like*. This is certainly problematic and should be remedied in future ELT coursebooks. Also, considering these findings, and taking into account Li's (2015) similar findings concerning the misrepresentation of the nouns *time* and *thing* in ELT coursebooks, it seems likely that other high-frequency words suffer from the same issues, which stem from a lack of awareness of the lexical syllabus. A major strength of this research, then, is that it can hopefully help in spreading awareness of this issue to ELT coursebook designers and to other researchers.

10.3.2: Limitations

I have identified three main limitations of the current study which need to be taken into consideration.

First, this study has investigated only one word, which limits its pedagogic applicability. While it is probable that the main pedagogic finding of this study is true for other high-frequency vocabulary items as well, it is not possible to say for certain without further research. Finding that *like* is not represented in ELT coursebooks in the same way as it is used in everyday English, further research will need to continue exploring this gap between corpora like the two I used.

The second limitation involves the corpora used. As with any corpus-based or corpus-driven study, the validity of the findings in the current study depends on the representativeness of the corpora used.

As I mentioned in Chapter 4, there are a few issues with the COCA. First, the spoken section is heavily weighted toward journalistic and news English with strong political overtones. Additionally, while the spoken section does consist of unscripted discourse, it can be criticized for not consisting of spontaneous and natural spoken English because the speakers were aware that they were being recorded. See Subsection 4.3.1 for further discussion of the COCA and its limitations. Next, during my data collection process, it became apparent that there are frequent data tagging errors in the COCA. This introduced the need for careful checking of each concordance line in the sorted data when compiling proportion and percentage figures. Finally, the monitor nature of the COCA, whereby we understand that it is periodically updated, can cause problems with consistency in the data for research conducted over a considerable length of time (e.g., differences in collocational MI scores).

With regard to the ELTCC, the main limitations are the limited size of the corpus and lack of part-of-speech (POS) tagging. Specialized corpora of over a million words are generally considered adequate for the type of comparative analysis conducted in this study. However, a much larger corpus would have allowed for a higher degree of confidence that the results obtained are fully representative of ELT coursebooks as a whole. Additionally, while the coursebooks in the ELTCC are all focused on general, communicative English, the corpus could be greatly improved with the inclusion of subcorpora, for example including coursebooks focused on written English, listening skills, and English for specific purposes. This would have provided the capacity for a much more thorough comparative analysis. The addition of POS tagging in the ELTCC would have been helpful for the comparative analyses, as the COCA is tagged for POS. Since the target word, *like*, is used across several parts of speech, it was on occasion time-consuming to manually check and sort data samples by POS. This was because I would need, for example, to manually remove all verbal uses of *like* from a prepositional *like* construction.

The third limitation of this study is concerned with the unavoidable issue of researcher subjectivity involved in the qualitative analysis of corpus data. There were many instances during the course of this research, especially while coding and sorting concordance data, where there was no clear choice between two or more options. This trouble would arise due to the data being part of a continuum, rather than a clear dichotomy. It proved problematic in two areas. One area lay in determining the specific sense of a polysemous item (see Nerlich et al. 2003 and Hanks 2013 for discussions on polysemy and subjectivity). And, similarly, the other area lay in determining whether a specific comparison was being used in a literal or figurative

manner. There were numerous instances in the data where it could be argued that a specific usage was from either of two senses of a construction or was both a literal usage and a figurative usage. In these cases, I subjectively chose to which category the data should be assigned, and it is certainly possible that there are instances where a different researcher may have chosen differently. This study could have benefited from having a second researcher to examine these problematic data in order to provide a sense of inter-rater reliability, and this will be considered in my future research.

10.4: Suggestions for Further Research

The results reported in this thesis (as discussed in Section 10.2) as well as the limitations of the study (discussed in Section 10.3) pave the way for three main areas that could benefit from further research.

First, considering the notable findings related to simile and genre with the *BE like* construction (and their implications for language learning) similar studies could be carried out with the simile types that were not covered in the current study. It has been noted that the majority of simile research is based on similes with *like*, as these are typically considered to be the prototypical simile type. Future studies could look at other simile types—for example, *as ADJECTIVE as*, and *as ADVERB as* similes—and compare findings to those of the current thesis. This would be helpful in determining if the same genre differences pertain to these other simile types as well. Future studies using the COCA could also take advantage of the three new genres (TV/movies, blogs, and general websites) that were added after the initial data had been collected for this study.

The second area for further research comes from one of the main limitations of this study: that only one word was investigated. Considering

the noteworthy results of this study, it would be greatly beneficial to continue the same line of research with other highly-frequent lexical items. Following the main principle of the lexical syllabus, this line of research should continue with the most frequent words in English. This could be approached by checking frequency lists for the top-listed content words that have yet to be analyzed. For example, looking at the COCA word frequency list we find several content words in the top-100 frequency band that could benefit from similar comparative corpora analyses, such as *about* (46th), *know* (47th), and *year* (54th), as well as the delexicalized verbs *get* (39th), *make* (45th), and *take* (63rd). Such research would aim to determine whether the problem—of *like* and its representation in ELT coursebooks with (at best) skewed regard for the principles of the lexical syllabus—applies for other words as well. It should also be noted that similar research has been carried out by Li (2015) on the words *time* (52nd) and *thing* (97th), in which the author came to a similar conclusion compared with that found in this study. These two high-frequency words were both underrepresented and misrepresented in ELT coursebooks.

Finally, classroom-based research could focus on teaching materials, designed to represent *like*, that follow the principles of the lexical syllabus. Classroom-based research could measure the effectiveness of such materials. For example, DDL exercises using *like* could be given to students, and they could be tested to evaluate their misuse and misunderstanding of *like*, as described in Subsection 9.5.1. It is possible that learners' misuse and misunderstanding could be reduced by presenting usages of *like* that better represent actual usage as reflected in the COCA. This would hopefully lead to the creation of more effective ELT coursebooks.

Appendix 1

Coursebooks Included in the ELTCC

Title	Abbreviation	Author/s	Publisher	Date
<i>American English File: Starter (2nd Edition)</i>	AmEngFileStart	Oxenden, Latham-Koenig, Boyle.	Oxford	2013
<i>American English File 1</i>	AmEngFile1	Oxenden, Latham-Koenig, Seligson.	Oxford	2008
<i>American English File 3</i>	AmEngFile3	Oxenden, Latham-Koenig.	Oxford	2008
<i>American English File 4</i>	AmeEngFile4	Oxenden, Latham-Koenig.	Oxford	2009
<i>Communication Strategies 3</i>	ComStrat3	Liu, David, Rizzon.	Cengage	2008
<i>New Cutting Edge Pre-Intermediate</i>	CutEdgePreInt	Cunningham, Moor.	Pearson	2005
<i>English Firsthand Access</i>	EngFirstHandAcc	Helgesen, Brown, Wiltshier.	Pearson	2009
<i>English Firsthand Success</i>	EngFirstHandSucc	Helgesen, Brown, Wiltshier.	Pearson	2009
<i>English Firsthand 1</i>	EngFirstHand1	Helgesen, Brown, Wiltshier.	Pearson	2009
<i>English Firsthand 2</i>	EngFirstHand2	Helgesen, Brown, Wiltshier.	Pearson	2009
<i>Four Corners 1</i>	FourCorners1	Richards, Bohlke.	Cambridge	2012
<i>Four Corners 2</i>	FourCorners2	Richards, Bohlke.	Cambridge	2012
<i>Four Corners 3</i>	FourCorners3	Richards, Bohlke.	Cambridge	2012
<i>Four Corners 4</i>	FourCorners4	Richards, Bohlke.	Cambridge	2012
<i>New Interchange 1</i>	NewInt1	Richards, Hull Proctor.	Cambridge	1997
<i>New Interchange 2</i>	NewInt2	Richards, Hull, Proctor.	Cambridge	1997
<i>Interchange 3 (3rd Edition)</i>	NewInt3	Richards, Hull, Proctor.	Cambridge	2005
<i>Nice Talking with You 1</i>	NiceTalkWithYou1	Kenny, Woo.	Cambridge	2011
<i>Nice Talking</i>	NiceTalkWithYou2	Kenny.	Cambridge	2012

Title	Abbreviation	Author/s	Publisher	Date
<i>with You 2</i>				
<i>Passages 1</i> (2 nd Edition)	Passages1	Richards, Sandy.	Cambridge	2008
<i>Passages 2</i>	Passages2	Richards, Sandy.	Cambridge	1998
<i>Side by Side 1</i> (3 rd Edition)	SideBySide1	Molinsky, Bliss.	Longman	2001
<i>Side by Side 2</i> (3 rd Edition)	SideBySide2	Molinsky, Bliss.	Longman	2001
<i>Side by Side 3</i>	SideBySide3	Molinsky, Bliss.	Longman	2001
<i>Touchstone 1</i>	Touchstone1	McCarthy, McCarten, Sandiford.	Cambridge	2005
<i>Touchstone 2</i>	Touchstone2	McCarthy, McCarten, Sandiford.	Cambridge	2005
<i>Touchstone 3</i>	Touchstone3	McCarthy, McCarten, Sandiford.	Cambridge	2005
<i>Talk Your Head Off</i> (And Write Too)	TalkYourHeadOff	West, West.	Pearson	1996
<i>World English 1</i>	WorldEng1	Milner.	Cengage	2010
<i>World English 2</i>	WorldEng2	Johannsen, Chase.	Cengage	2010
<i>World English 3</i> (2 nd Edition)	WorldEng3	Johannsen, Chase.	Cengage	2010
<i>World Link Intro</i>	WorldLinkInt	Stempleski, Douglas, Morgan.	Cengage	2004
<i>World Link 1</i>	WorldLink1	Stempleski, Douglas, Morgan.	Cengage	2004
<i>World Link 2</i>	WorldLink2	Stempleski, Douglas, Morgan.	Cengage	2004

Appendix 2

Simile: Categories of Targets and Vehicles

During the initial sorting and ad hoc categorization of the *NOUN BE like NOUN* data samples, I identified a total of 27 unique categories of targets and vehicles, as follows:

1. Organization
2. People
3. Business
4. Behaviour
5. Event/state
6. Physical object
7. Field of study/practice
8. Abstract thing
9. Cognitive/feeling
10. Supernatural/character
11. Body part/s
12. Sickness/disease
13. Nature
14. Drug/medicine
15. Pastime/hobby/sport
16. Place
17. Animal
18. Government
19. Action
20. Food
21. Building
22. Technology
23. Verbal
24. Sound/music
25. Sensation (touch, sight etc.)
26. Media
27. Time

References

- Adolphs, S., and R. Carter. 2003. "And She's like It's Terrible, like: Spoken Discourse, Grammar and Corpus Analysis." *International Journal of English Studies* 3: 45–56.
- Aisenman, R. A. 1999. "Structure-Mapping and the Simile-Metaphor Preference." *Metaphor and Symbol* 14 (1): 45–51.
- Anthony, L. 2015. TagAnt (Version 1.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Anthony, L. 2018. AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Aristotle. 1984. *The Complete Works of Aristotle: The Revised Oxford Translation*. Edited by Jonathan Barnes. 2 vols. Bollingen Series. Princeton, NJ: Princeton University Press.
- Bauer, L. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Biber, D. 1999. "Introduction: A Corpus-Based Approach to English Grammar." *Longman Grammar of Spoken and Written English*, 4–5.
- Biber, D., S. Conrad, and V. Cortes. 2004. "If You Look at . . . : Lexical Bundles in University Teaching and Textbooks." *Applied Linguistics* 25 (3): 371–405.
- Biber, D., and R. Reppen. 2015. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Boulton, A. 2007a. "But where's the proof? The need for empirical evidence for data-driven learning." *40th Annual Meeting of the British Association for Applied Linguistics: Technology, Ideology and Practice in Applied Linguistics*. University of Edinburgh, Scotland (September 6–8).
- Boulton, A. 2007b. "DDL is in the details... and in the big themes." In Davies, M., Rayson, P., Hunston, S., and Danielsson, P. (eds.) *Proceedings of Corpus Linguistics 2007* (www) <http://www.corpus.bham.ac.uk/corplingproceedings07/> (January 8, 2010).
- Boulton, A. 2008. "DDL: Reaching the parts other teaching can't reach?" Ifrankenburg-Garcia, A. (ed) *Proceedings of the 8th teaching and language corpora conference*. Lisbon, Portugal: Associação de Estudos e de Investigação Científica do ISLA-Lisboa (July 3–6).
- Bowdle, B. F., and D. Gentner. 2005. "The Career of Metaphor." *Psychological Review* 112 (1): 193–216.
- Bredin, H. 1998. "Comparisons and Similes." *Lingua* 105 (97): 67–78.
- Bridgeman, T. 1996. "On the 'Likeness' of Similes and Metaphors (With Special Reference to Alfred Jarry's *Les Jours et Les Nuits*)." *The Modern Language Review* 91 (1): 65–77.
- Browne, C., B. Culligan, and J. Phillips. 2013. "New Academic Word List 1.0." <http://www.newgeneralservicelist.org/nawl-new-academic-word-list>.

- Browne, C., B. Culligan, and J. Phillips. 2017. "New General Service List: Spoken." <http://www.newgeneralservicelist.org/ngsls>.
- Burck, G. 1965. *The Computer Age and Its Potential for Management*. Harper & Row. P. 62.
- Burton, G. 2012. "Corpora and Coursebooks: Destined to Be Strangers Forever?" *Corpora* 7 (1).
- Cameron, L. and Deignan, A. 2003. "Combining large and small corpora to investigate tuning devices around metaphor in spoken discourse." *Metaphor and Symbol*, 18 (3): 149–160.
- Carter, R. 1987. *Vocabulary: Applied Linguistic Perspectives*. London, England: Allen and Unwin.
- Carter, R. 2007. "Spoken English, Written English: Challenging Assumptions" The 33rd JALT International Conference on Language Teaching and Learning. The Japan Association for Language Teaching, National Olympics Memorial Youth Center, Tokyo, Japan. (November 22–25, 2007).
- Carter, R., and M. McCarthy. 1995. "Grammar and the Spoken Language." *Applied Linguistics* 16.2: 141–58.
- Chalker, S. 1994. "Pedagogical Grammar: Principles and Problems," in Bygate, M.; Tonkyn, A.; and Williams, E. (eds.) *Grammar and the Language Teacher*. Hertfordshire: Prentice Hall.
- Chiappe, D. L., and J. M. Kennedy. 1999. "Aptness Predicts Preference for Metaphors or Similes, As Well As Recall Bias." *Psychonomic Bulletin and Review* 6 (4): 668–76.
- Chiappe, D. L., John M. Kennedy, and P. Chiappe. 2003. "Aptness Is More Important Than Comprehensibility in Preference for Metaphors and Similes." *Poetics* 31 (1): 51–68.
- Church, K. and Hanks, P. 1990. "Word association norms, mutual information, and lexicography." *Computational Linguistics* 16 (1): 22–29.
- Church, K., W. Gale, P. Hanks, and D. Hindle. 1991. "Using Statistics in Lexical Analysis" in U. Zernik (ed.), *Lexical Acquisition: Using On-Line Resources to Build a Lexicon*. Englewood Cliff, New Jersey: Lawrence Erlbaum.
- Cobb, T. and Boulton, A. "Classroom applications of corpus analysis." In *Cambridge Handbook of English Corpus Linguistics*, 478–497. Cambridge: Cambridge University Press.
- Collins COBUILD Learner's Dictionary. Concise Ed, 2003, s.v. "just," "like." Glasgow: HarperCollins.
- Conrad, J. 2007. *Heart of Darkness*. Edited by R. Hampson and O. Knowles. London, England: Penguin Classics.
- Cowie, A. P. 1981. "The Treatment of Collocations and Idioms in Learners' Dictionaries." *Applied Linguistics* 2 (3): 223–235.
- Cowie, A. P. 1994. "Phraseology." In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*, 3168–3171. Oxford: Oxford University Press.

- Cowie, A. P. 1998. *Phraseology: Theory, Analysis, and Applications*. Language.
- Criado, R. 2013. "A critical review of the Presentation-Practice-Production Model (PPP) in foreign language teaching." In R. Monroy (ed.), *Homenaje a Francisco Gutiérrez Díez*. Murcia: Edit.um.
- Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. New York: Oxford University Press.
- D'Arcy, A. 2006. "Lexical Replacement and the Like(S)." *American Speech* 81 (4): 339–357.
- D'Arcy, A. 2007. "Like and Language Ideology: Disentangling Fact from Fiction." *American Speech* 82 (4): 386–419.
- D'Arcy, A. 2017. *Discourse-Pragmatic Variation in Context: Eight Hundred Years of LIKE*. Amsterdam: John Benjamins.
- Davies, M. 2020. Corpus of Contemporary American English (website). Updated March 2020. <https://www.english-corpora.org/coca/>.
- De Mey, M. 1982. *The Cognitive Paradigm*. London, England: D. Reidel Publishing.
- Dictionary.com. Accessed September 26, 2020. <https://www.dictionary.com>.
- Downing, A., and P. Locke. 2006. *English Grammar: A University Course*. 2nd ed. Oxon: Routledge.
- Durrant, P. and Schmitt, N. 2009. "To what extent do native and non-native writers make use of collocations?" *International Review of Applied Linguistics in Language Teaching* 47: 157–177.
- Fadaee, E. 2011. "Symbols, Metaphors and Similes in Literature: A Case Study of Animal Farm" *Journal of English and Literature* 2 (2): 19–27.
- Fillmore, C., P. Kay, and M. C. O'Connor 1988. "Regularity and Idiomaticity in Grammatical Constructions." *Language*, 64: 501–538.
- Firth, J. R. 1957. "A Synopsis of Linguistic Theory, 1930–1955." *Studies in Linguistic Analysis* Special Volume, Philological Society. 1–32.
- Fishelov, D. 1993. "Poetic and Non-Poetic Simile: Structure, Semantics, Rhetoric." *Poetics Today* 14 (1): 1–23.
- Francis, G., and J. Sinclair. 1994. "'I Bet He Drinks Carling Black Label:' A Riposte to Owen on Corpus Grammar." *Applied Linguistics* 15 (2): 190–200.
- Fromilhague, C. 1995. *Les Figures de Style*. Paris: Nathan.
- Gavioli, L. 2005. *Exploring Corpora for ESP Learning*. Philadelphia: John Benjamins.
- Gentner, D., and B. F. Bowdle. 2001. "Convention, Form, and Figurative Language Processing." *Metaphor and Symbol* 16 (3–4): 223–47.
- Gibbs, R. W., Jr. 1994. *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.
- Gläser, R. 1998. "The Stylistic Potential of Phraseological Units in the Light of Genre Analysis." In A.P. Cowie (ed.). *Phraseology: Theory, Analysis, and Applications*, 125–143. Oxford: Oxford University Press.

- Glucksberg, S. 1980. "The Psycholinguistics of Metaphor." *Trends in Cognitive Sciences* 7 (2): 92–96.
- Glucksberg, S., and C. Haught. 2006a. "Can Florida Become *Like* the Next Florida?" *Psychological Science* 17 (11): 935–38.
- Glucksberg, S., and C. Haught. 2006b. "On the Relation Between Metaphor and Simile: When Comparison Fails." *Mind and Language* 21 (3): 360–78.
- Glucksberg, S., and B. Keysar. 1990. "Understanding Metaphorical Comparisons: Beyond Similarity." *Psychological Review* 97 (1): 3–18.
- Glucksberg, S., and M. S. McGlone. 2001. *Understanding Figurative Language : From Metaphors to Idioms*. New York: Oxford University Press.
- Goatly, A. 1997. *The Language of Metaphors*. New York: Routledge.
- Gokcesu, B. S. 2009. "Comparison, Categorization, and Metaphor Comprehension." *Proceedings of the Annual Meeting of the Cognitive Science Society* 31 (31): 567–72.
- Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Go Natural English (website). Accessed September 26, 2020. "1000 Most Common Words in English." <https://www.gonaturalenglish.com/1000-most-common-words-in-the-english-language/>.
- Granger, S., and F. Meunier, eds. 2008. *Phraseology: An Interdisciplinary Perspective*. Philadelphia: John Benjamins.
- Granger, S., and M. Paquot. 2008. "Disentangling the Phraseological Web." In *Phraseology: An Interdisciplinary Perspective*, edited by S. Granger and F. Meunier, 27–49. Amsterdam: John Benjamins.
- Gray, B., and D. Biber. 2015. "Phraseology." In *The Cambridge Handbook of English Corpus Linguistics*, edited by D. Biber and R. Reppen, 125–45. Cambridge: Cambridge University Press.
- Greenbaum, S. 1974. "Some Verb-Intensifier Collocations in American and British English." *American Speech* 49 (1): 79–89.
- Gries, S. T. 2008. "Phraseology and Linguistic Theory: A Brief Survey." In *Phraseology: An Interdisciplinary Perspective*, edited by S. Granger and F. Meunier, 3–25. Amsterdam: John Benjamins. 3–25.
- Gries, S. T. 2015. "Statistics for Learner Corpus Research." In *The Cambridge Handbook of Learner Corpus Research*, edited by S. Granger, G. Gilquin, and F. Meunier, 159–182. Cambridge: Cambridge University Press.
- Groom, N. 2017. "Phraseology: A Critical Reassessment." Paper given at the Idiomaticity Workshop, University of Oslo, Norway, 2 September 2017.
- Groom, N. 2019. "Construction Grammar and the Corpus-Based Analysis of Discourses The Case of the WAY IN WHICH Construction." *International Journal of Corpus Linguistics* 24 (3): 291–323.

- Halliday, M. A. K. 1961. "Categories of the Theory of Grammar." *WORD* 17 (2): 241–92.
- Halliday, M. A. K. 1976. *Learning How to Mean: Explorations in the Development of Language*. London: Edward Arnold.
- Halliday, M. A. K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, M. A. K., and Matthiessen, C. M. I. M. 2004. *An Introduction to Functional Grammar* (3rd ed.). London: Edward Arnold.
- Hanks, P. 2005. "Similes and Sets: The English Preposition *like*." In *Jazyky a Jazykověda (Languages and Linguistics: Festschrift for Professor Fr. Čermák)*, edited by R. Blatná and V. Petkevič. Prague: Charles University.
- Hanks, P. 2013. *Lexical analysis: Norms and Exploitations*. Cambridge: MIT Press.
- Hasan, R. 1996. *Ways of Saying, Ways of Meaning: Selected Papers*, edited by C. Cloran, D. Gutt, and G. Williams, 73–103. London: Cassell.
- Herron, C. and Tomasello, M. 1992 "Acquiring grammatical structures by guided induction." *The French Review* 65 (5): 708–718.
- Hilpert, M. 2014. *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.
- Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Howarth, P. A. 1996. *Phraseology in English Academic Writing*. Tübingen: Max Niemeyer Verlag.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. 2003. "Lexis, Wordform and Complementation Pattern: A Corpus Study." *Functions of Language*, 10, 31–60.
- Hunston, S. 2010. *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London: Routledge.
- Hunston, S., and G. Francis. 1998. "Verbs Observed : A Corpus-Driven Pedagogic Grammar." *Applied Linguistics* 19 (1): 45–72.
- Hunston, S., and G. Francis. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English. Language*. Amsterdam: John Benjamins.
- Hunston, S., and S. Laviosa. 2001. *Corpus Linguistics*. Birmingham: University of Birmingham Centre for English Language Studies.
- Hunston, S., and H. Su. 2017. "Patterns, Constructions, and Local Grammar: A Case Study of 'Evaluation'". *Applied Linguistics* 40 (4): 1–28.
- Imao, Y. 2018. CasualConc (Version 2.1.1) [Computer Software]. Osaka, Japan: Osaka University. Available from: <https://sites.google.com/site/casualconc/>
- Nunan, D. 1988. *Syllabus Design*. Oxford: Oxford University Press.
- Johns, T. 1991. "Should You Be Persuaded: Two Examples of Data-Driven Learning." *ELR Journal: Classroom Concordancing* 4 (Farrington 1989): 1–16.

- Johns, T. 1994. "From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-Driven Learning." In *Perspectives on Pedagogical Grammar*, edited by T. Odlin, 293–313. Cambridge: Cambridge University Press.
- Johnson, M. G., and R. G. Malgady. 1979. "Some Cognitive Aspects of Figurative Language: Association and Metaphor." *Journal of Psycholinguistic Research* 8 (3): 249–65.
- Jones, L. L., and Z. Estes. 2005. "Metaphor Comprehension as Attributive Categorization." *Journal of Memory and Language* 53 (1): 110–24.
- Jones, L. L., and Z. Estes. 2006. "Roosters, Robins, and Alarm Clocks: Aptness and Conventionality in Metaphor Comprehension." *Journal of Memory and Language* 55 (1): 18–32.
- Kennedy, G. 1991. "Between and Through: The Company They Keep and the Functions They Serve." in *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, edited by K. Aijmer and B. Altenberg, 95–110. London: Longman.
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Kisak, P. F. 2015. *Literary Devices: All Writers Should Know!* Createspace Indep Pub.
- Koprowski, M. 2005. "Investigating the Usefulness of Lexical Phrases in Contemporary Coursebooks." *ELT Journal* 59 (October): 322–32.
- Lakoff, G., and M. Johnson, 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lakoff, G., and M. Johnson. 2003. "Afterword, 2003." In *Metaphors We Live By*. 2nd ed. Chicago: The University of Chicago Press.
- Lebovits, G. 2002. "What's Another Word for 'Synonym'?" *New York State Bar Journal*, 2002 - works.bepress.com
- Lewis, M. 1993. *The Lexical Approach*. Tokyo: Language Teaching Publications.
- Lewis, M. 1996. "Implications of a Lexical View of Language." In *Challenge and Change in Language Teaching*, edited by J. Willis and D. Willis. Oxford: Macmillan.
- Li, S. 2015. "A Corpus-Based Study of the High Frequency Nouns Time and thing: Investigating the Role of Phraseology in the Construction of Meaning in Discourse." PhD thesis, University of Birmingham.
- Lightbrown, P., and N. Spada. 2006. *How Languages Are Learned*. 3rd ed. Oxford: Oxford University Press.
- Littlejohn, A. P. 1992. "Why Are English Language Teaching Materials the Way They Are?" PhD thesis, Lancaster University.
- Littlemore, J., and Low, G. 2006. *Figurative Thinking and Foreign Language Learning*. New York: Palgrave Macmillan.
- Linguamarina. "100 Most Common English Words: Beginner Vocabulary," published December 31, 2017. YouTube video. <https://www.youtube.com/watch?v=oOZLPVG9-9I>.
- Margolis, J. 1957. "Notes on the Logic of Simile, Metaphor and Analogy Joseph Margolis." *American Speech* 32 (3): 186–89.

- Martel, J. 2013. "Saying our final goodbyes to the grammatical syllabus: A curricular imperative." *The French Review* 86 (6): 1122-1133.
- Mason, O. 1997. "The Weight of Words: An Investigation of Lexical Gravity." In *PALC'97: Practical Applications in Language Corpora*, edited by B. Lewandowska-Tomaszczyk and P. J. Melia, 361-75. Lodz: Lodz University.
- Mason, O. 1999. "Parameters of Collocation: The Word in the Centre of Gravity." In *Corpora Galore: Analyses and Techniques in Describing English*, edited by J. Kirk, 267-280. Amsterdam: Rodopi.
- McCarthy, M. 1990. *Vocabulary*. Oxford: Oxford University Press.
- McClean, S. 2018. "Evidence for the adoption of the flemma as an appropriate word counting unit." *Applied Linguistics* 39 (8): 823-845.
- McGlone, M. S. 2007. "What Is the Explanatory Value of a Conceptual Metaphor?" *Language and Communication* 27 (2): 109-26.
- McGuigan, Brendan. 2007. *Rhetorical Devices: A Handbook and Activities for Student Writers*. Clayton, Delaware: Prestwick House.
- Mel'čuk, I. A. 1998. "Collocations and Lexical Functions." In *Phraseology: Theory, Analysis and Applications*, edited by A. P. Cowie, 23-53. Oxford: Oxford University Press.
- Mel'čuk, I. A. 2012. "Semantics: From Meaning to Text." Amsterdam: John Benjamins. Pp. xxi, 436.
- Merriam-Webster*, s.vv. "just," "like," accessed September 26, 2020, <https://www.merriam-webster.com/dictionary/just>, <https://www.merriam-webster.com/dictionary/like>.
- Miller, G. A. 1993. "Images and Models, Similes and Metaphors." In *Metaphor and Thought*, edited by Andrew Ortony, 357-400. Cambridge: Cambridge University Press.
- Milner, M. 2012. *World English 1*. Boston: Cengage.
- Moon, R. 1998. *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- Moon, R. 2011a. "English Adjectives in *-like*, and the Interplay of Collocation and Morphology." *International Journal of Corpus Linguistics* 16 (4): 486-513.
- Moon, R. 2011b. "Simile and Dissimilarity." *Journal of Literary Semantics*. 40, 2, 133-157.
- Nattinger, J. R., and J. S. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Neal, A. 2006. "'Matching' Corpus Data and System Networks: Using Corpora to Modify and Extend the System Networks for TRANSITIVITY in English." In *System and Corpus: Exploring Connections*, edited by G. Thompson and S. Hunston. London: Equinox.
- Nesselhauf, N. 2004. "What are collocations?" In *Phraseological Units: Basic Concepts and Their Application*, edited by D. J. Allerton, N. Nesselhauf, and P. Skandera, 1-21. Basel: Schwabe.
- Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Nunan, D. 1988. *Syllabus Design*. Oxford: Oxford University Press.

- O'Keeffe, A., M. McCarthy, and R. Carter. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. New York: Cambridge University Press.
- Ortony, Andrew. 1979. "Beyond Literal Similarity." *Psychological Review* 86 (3): 161–80.
- Oxford English Dictionary*, s.v. "like," accessed September 26, 2020. <https://www.oed.com>.
- Pawley, A., and F. H. Syder. 1983. "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency." *Language and Communication* 5 (5): 191–226.
- Peppard, J. 2014. "A Lexicogrammatical Approach to Fluency." In *Exploring EFL Fluency in Asia*, edited by T. Muller, J. Adamson, P. S. Brown, and S. Herder. New York: Palgrave Macmillan.
- Peppard, J. 2016. "Like That? The Problem with Coursebook Vocabulary." In *JALT 2015: Focus on the Learner*, edited by P. Clements, A. Krause, and H. Brown, 196–97. https://jalt-publications.org/proceedings/issues/2016-08_2015.1.
- Pinker, S. 1994. *The Language Instinct*. New York: HarperCollins.
- Renouf, A., and J. Sinclair. 1991. "Collocational Frameworks in English." *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, 128–43.
- Richards, I. A. 1965. *The Philosophy of Rhetoric*. Oxford University Press.
- Richards, J.C., and D. Bohlke. 2012. *Four Corners: Student's Book 4*. Cambridge: Cambridge University Press.
- Richards, J. C., and R. W. Schmidt. 2002. *Longman Dictionary of Language Teaching and Applied Linguistics*. London: Longman.
- Romer, U. 2006. "Pedagogical Applications of Corpora: Some Reflections on the Current Scope and a Wish List for Future Developments." *Zeitschrift für Anglistik und Amerikanistik* 54 (2): 121–134.
- Roncero, C., J. M. Kennedy, and R. Smyth. 2006. "Similes on the Internet Have Explanations." *Psychonomic Bulletin and Review* 13 (1): 74–77.
- Rutherford, W. 1987. *Second Language Grammar: Learning and Teaching*. New York: Pearson.
- Sharwood-Smith, M. 1981. "Consciousness-raising and the second language learner." *Applied Linguistics* 2: 159–68.
- Sherry, J. F. 2007. "Fruit Flies Like a Banana (Or, When Ripeness Is All): A Meditation on Markets and Timescapes." In *Multi-Level Issues in Organizations and Time*, edited by F. Dansereau and F. J. Yammarino, F.J. (Ed.), 331–337. Bingley, West Yorkshire: Emerald Group Publishing.
- Shibata, M. 2012. "What Is the Difference Between Metaphor and Simile? fMRI Study; Midori." *CARLS Series of Advanced Study of Logic and Sensibility* 5: 101–9.
- Shibata, M., A. Toyomura, H. Motoyama, H. Itoh, Y. Kawabata, and J. Abe. 2012. "Does Simile Comprehension Differ from Metaphor Comprehension? A Functional MRI Study." *Brain and Language* 121 (3): 254–260.
- Shortall, T. 2007. "The L2 Syllabus: Corpus or Contrivance?" *Corpora* 2 (2): 157–85.

- Sinclair, J. 1966. "Beginning the Study of Lexis." In *In memory of J. R. Firth*, edited by C. E. Bazell, J. C. Catford, H. A. K. Halliday, and R. H. Robins, 410–430. London: Longmans.
- Sinclair, J. 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair, J. 1998. "The Lexical Item." In *Contrastive lexical semantics*, edited by E. Weigand. Amsterdam: John Benjamins.
- Sinclair, J. 2004. "The Lexical Item." In *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Sinclair, J. 2005. Postgraduate seminar held at the University of Birmingham, May 2005.
- Sinclair, J., and S. Jones. 1974. "English Lexical Collocations: A Study in Computational Linguistics." In *J. M. Sinclair on Lexis and Lexicography*, edited by J. Foley. Singapore: University of Singapore Press.
- Sinclair, J. M., and A. Renouf. 1988. "A Lexical Syllabus for Language Learning." In *Vocabulary and Language Teaching*, edited by R. Carter and M. McCarthy, 140–60. Oxon: Routledge.
- Sinclair, J. M. 2003. *Reading Concordances: An Introduction*. London: Longman.
- Skehan, P. 1996. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Steen, G. 2008. "The Paradox of Metaphor: Why We Need a Three-Dimensional Model of Metaphor." *Metaphor and Symbol* 23 (4): 213–41.
- Stubbs, M. 1996. *Text and Corpus Analysis*. Oxford: Blackwell Publishing.
- Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishing.
- Stubbs, M. 2002. "Two Quantitative Methods of Studying Phraseology in English." *International Journal of Corpus Linguistics* 7 (2): 215–44.
- Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.
- Taylor, J. 2012. *The Mental Corpus: How Language is Represented in the Mind*. Oxford: Oxford University Press.
- Taylor, L. 1991. "Review: *Collins COBUILD English Course*." *ELT Journal* 45 (1): 74–77.
- Tian, S. P. 2004. "Data-Driven Learning: Do Learning Tasks and Proficiency Make a Difference?" Paper presented at the 9th Pan-Pacific Association of Applied Linguistics (PAAL) Conference, Namseoul University, Korea, August 2004.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tucker, G. 2006. "Systemic Incorporation: On the Relationship Between Corpus and Systemic Functional Grammar." In *System and Corpus: Exploring Connections*, Edited by G. Thompson and S. Hunston, 81–102. London: Equinox.
- van Ek, J. 1975. *The Threshold Level English*. Pergamon: Oxford University Press.
- Veale, T., and Y. Hao. 2007. "Learning to Understand Figurative Language: From Similes to Metaphors to Irony." In *Proceedings of CogSci 2007, the 29th Annual Meeting of the Cognitive Science Society*. New York: Lawrence Erlbaum Associates.

- Viberg, Å. 1983. "The Verbs of Perception: A Typological Study." *Linguistics* 21 (1): 123–62.
- Walker, C. 2008a. "A Corpus-Based Study of the Linguistic Features and Processes Which Influence the Way Collocations Are Formed." PhD thesis, University of Birmingham.
- Walker, C. 2008b. "Factors Which Influence the Process of Collocation." In *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*, edited by F. Boers and S. Lindstromberg, 291–308. Berlin: Mouton de Gruyter.
- Walker, C. P. 2011. "A Corpus-Based Study of the Linguistic Features and Processes Which Influence the Way Collocations Are Formed: Some Implications for the Learning of Collocations." *TESOL Quarterly* 45 (2): 291–312.
- Wikberg, K. 2008. "Phrasal Similes in the BNC." In *Phraseology: An Interdisciplinary Perspective*, edited by S. Granger and F. Meunier, 127–142. Philadelphia: John Benjamins.
- Wiktionary (website). "Frequency Lists/TV/2006/1–1000," last edited November 10, 2019. https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/TV/2006/1-1000.
- Willis, D. 1990. *The Lexical Syllabus: A New Approach to Language Teaching*. London: Collins ELT.
- Willis, J. 1996. "A flexible framework for task-based learning." In *Challenge and Change in Language Teaching*, edited by J. Willis and D. Willis.
- Willis, D. 2003. *Rules, Patterns and Words: Grammar and Lexis in English Language Teaching*. Cambridge: Cambridge University Press.
- Willis, J., and D. Willis. 1988. *Collins COBUILD English Course: Student's Book 1*. London: Collins.
- Winter, B. 2019. *Sensory Linguistics: Language, Perception and Metaphor*. Amsterdam: John Benjamins.
- Wray, A. 2000. "Formulaic Sequences in Second Language Teaching: Principle and Practice." *Applied Linguistics* 21 (4): 463–89.
- Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wu, L. 2016. "An Analysis of Japanese English Learners' Errors." *Bulletin of Yamagata University (Educational Science)* 16 (3): 35–47.