

ROUTLEDGE ADVANCES IN CORPUS LINGUISTICS


Corpus Stylistics and Dickens's Fiction

Michaela Mahlberg

Corpus Stylistics and Dickens's Fiction

Michaela Mahlberg



 **Routledge**
Taylor & Francis Group
NEW YORK AND LONDON

First published 2013
by Routledge
711 Third Avenue, New York, NY 10017

Simultaneously published in the UK
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

*Routledge is an imprint of the Taylor & Francis Group,
an informa business*

© 2013 Taylor & Francis

The right of Michaela Mahlberg to be identified as author of this work
has been asserted by her in accordance with sections 77 and 78 of the
Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or
utilised in any form or by any electronic, mechanical, or other means, now
known or hereafter invented, including photocopying and recording, or in
any information storage or retrieval system, without permission in writing
from the publishers.

Trademark Notice: Product or corporate names may be trademarks or
registered trademarks, and are used only for identification and explanation
without intent to infringe.

Library of Congress Cataloging in Publication Data

Mahlberg, Michaela.

Corpus stylistics and Dickens's fiction / by Michaela Mahlberg.

p. cm.—(Routledge advances in corpus linguistics; 14)

Includes bibliographical references and index.

ISBN 978-0-415-80014-3

1. Corpora (Linguistics) 2. English language—Style.
 3. Discourse analysis, Literary. 4. Dickens, Charles, 1812–1870—
Language. 5. Dickens, Charles, 1812–1870—Criticism, Textual.
- I. Title.

P128.C68M34 2012

823'.8—dc23

2012028000

ISBN: 978-0-415-80014-3 (hbk)

ISBN: 978-0-203-07608-8 (ebk)

Typeset in Sabon
by Apex CoVantage, LLC

“Michaela Mahlberg’s work on Dickens is ground-breaking and genuinely important because it analyses and brings into dialogue major aspects of his writing that have been wrongly and strangely neglected in modern criticism: in particular, his language and characterisation. The neglect of these fundamental characteristics of Dickens’s fiction may have resulted from disciplinary fissures, the divergence of ‘Language’ from ‘Literature’, History from Aesthetics. Here too, Mahlberg proves herself a pioneer, not only integrating the methods of English Language—in particular, Corpus Linguistics—with English Literature, but in so doing, applying technology creatively to textual analysis. *Corpus Stylistics and Dickens’s Fiction* opens up a whole new vista on Dickens’s work, suggesting fertile and productive angles for future research.”

—Juliet John, Royal Holloway,
University of London, UK

“Mahlberg offers us a first-rate study of Dickens’s fiction, carried out at the intersection of four major domains of research: corpus-linguistic research methods, the study of phraseology, text linguistics, and literary stylistics. By applying corpus-driven analytical methods, Mahlberg helps us understand some of the major phraseological devices that Dickens employs to create the memorable characters that we all love. This book will be of interest for both its theoretical contributions in the areas of corpus linguistics and literary stylistics, as well as for its detailed descriptions of Dickens’s style.”

—Douglas Biber,
Northern Arizona University, USA

“Michaela Mahlberg’s book provides both a structured introduction to the field of corpus stylistics, which has become increasingly popular over the past few years, and corpus-stylistic case studies focusing on Charles Dickens’s work. Mahlberg shows that the combination of corpus-linguistic methods and literary-stylistic approaches opens up entirely new perspectives on the description and analysis of how literary texts ‘work’. This book shows in a refreshingly innovative way how the great divide between literary preoccupation with the creative potential of individual authors and texts on the one hand and the corpus-linguistic identification of general routines in language use on the other can be overcome—this book is a must for everyone who intends to join the ever-growing corpus-stylistic community.”

—Joybrato Mukherjee,
University of Giessen, Germany

"I've been eagerly awaiting the publication of this innovative book for some months now. Michaela Mahlberg's work is both highly creative and rigorous methodologically. I think it is a must-read book for anyone interested in corpus stylistics, stylistics more generally, corpus linguistics and Dickens studies. Mahlberg reveals fascinating patterns of 'phrasal' repetition scattered through Dickens's novels which influence readers, mainly subliminally, when they read his work."

—Mick Short,
Lancaster University, UK

Corpus Stylistics and Dickens's Fiction

This book presents an innovative approach to the language of one of the most popular English authors. It illustrates how corpus linguistic methods can be employed to study electronic versions of texts by Charles Dickens. With particular focus on Dickens's novels, the book proposes a way into the Dickensian world that starts from linguistic patterns. The analysis begins with clusters, i.e., repeated sequences of words, as pointers to local textual functions. Combining quantitative findings with qualitative analyses, the book takes a fresh view on Dickens's techniques of characterisation, the literary presentation of body language and speech in fiction. The approach brings together corpus linguistics, literary stylistics, and Dickens criticism. It thus contributes to bridging the gap between linguistic and literary studies and will be a useful resource for both researchers and students of English language and literature.

Michaela Mahlberg is associate professor in English language and applied linguistics at the University of Nottingham, UK.

Routledge Advances in Corpus Linguistics

Edited by Tony McEnery, *Lancaster University, UK*

Michael Hoey, *Liverpool University, UK*

- 1 Swearing in English**
Bad Language, Purity and Power
from 1586 to the Present
Tony McEnery
- 2 Antonymy**
A Corpus-Based Perspective
Steven Jones
- 3 Modelling Variation in Spoken
and Written English**
David Y. W. Lee
- 4 The Linguistics of Political
Argument**
The Spin-Doctor and the Wolf-Pack
at the White House
Alan Partington
- 5 Corpus Stylistics**
Speech, Writing and Thought
Presentation in a Corpus of
English Writing
Elena Semino and Mick Short
- 6 Discourse Markers Across
Languages**
A Contrastive Study of Second-
Level Discourse Markers in
Native and Non-Native Text with
Implications for General and
Pedagogic Lexicography
Dirk Siepmann
- 7 Grammaticalization and English
Complex Prepositions**
A Corpus-Based Study
Sebastian Hoffmann
- 8 Public Discourses of Gay Men**
Paul Baker
- 9 Semantic Prosody**
A Critical Evaluation
Dominic Stewart
- 10 Corpus Assisted Discourse
Studies on the Iraq Conflict**
Wording the War
*Edited by John Morley and Paul
Bailey*
- 11 Corpus-Based Contrastive
Studies of English and Chinese**
Richard Xiao and Tony McEnery
- 12 The Discourse of Teaching
Practice Feedback**
A Corpus-Based Investigation of
Spoken and Written Modes
Fiona Farr
- 13 Corpus Approaches to Evaluation**
Susan Hunston
- 14 Corpus Stylistics and Dickens's
Fiction**
Michaela Mahlberg

Acknowledgements

On the journey to the completion of this book I was very fortunate to have the support of many friends and colleagues. I would like to thank in particular Peter Stockwell, Ron Carter, Beatrix Busse, Fanie Tsiamita, Anna Čermáková, Kathy Conklin, and an anonymous reviewer for commenting on the draft manuscript of this book or parts of it. I am especially indebted to Catherine Smith, with whom I have worked on publications related to the concerns of this book. She has also provided the technology to retrieve the examples discussed in Chapter 9. I am grateful for inspiring conversations with Mick Short, Geoffrey Leech, Michael Toolan, Dan McIntyre, Doug Biber, Mike Scott, Juliet John, Josephine Guy, Mark Llewellyn, and the late John Sinclair that influenced and helped shape the content of this book. My students at the universities of Liverpool and Nottingham were confronted with early ideas of what made it into this book in some form or other; and I am particularly grateful to Katherine Gupta and Lorenzo Mastropierro for their feedback. I thank Tony McEnery and Michael Hoey who encouraged and supported this project as series editors, and Felisa Salvago-Keyes at Routledge for her patience and cooperation. My thanks also go the School of English at the University of Nottingham for funding the final formatting and editing work.

This book also greatly benefited from the feedback and constructive criticism that I received when presenting early versions of work developed in these chapters. I would like to thank in particular the following people for inviting me to speak at conferences or other events: Paul Baker, Michael Burke, Pernilla Danielsson, Izaskun Elorza, Maristella Gatto, Lesley Jeffries, Ramesh Krishnamurthy, Dilin Liu, Joybrato Mukherjee, Lynne Murphy, Patricia Ronan, Mick Short, Nick Smith, Wolfgang Teubert, and Martin Wynne.

My greatest thanks go to Heiko Schilling, my husband, our son Richard, and my parents Elisabeth and Fritz Mahlberg—this book truly was a team effort in many ways.

Michaela Mahlberg
Nottingham, May 2012

The author and publisher are also grateful to the following for permission to reprint copyright material:

Edinburgh University Press (www.euppublishing.com) for permission to include an adapted version of Figure 4.2 from J. Gavins (2007). *Text World Theory. An Introduction*, p. 58, © J. Gavins 2007 (see Figure 2.1).

University of Chicago Press for permission to include the transcribed notes for chapter 1 of *Little Dorrit* from H. Stone (1987). *Dickens' Working Notes for his Novels*, p. 271, © 1987 by The University of Chicago (see Figure 9.2).

Contents

<i>List of Tables, Figures, and Concordances</i>	xi
<i>Acknowledgements</i>	xiii
Introduction	1
1 Corpus Stylistics	5
Introduction	5
1.1 Shared ground and innovation	7
1.1.1 Foregrounding theory and corpus norms	8
1.1.2 The philological and the corpus stylistic circle	11
1.1.3 The stylistician's toolkit and units of description	14
1.2 Examples of corpus stylistic work	18
1.3 Entering the corpus stylistic circle for the study of Dickens's fiction	23
Conclusions	24
2 Textual Building Blocks of Fictional Worlds	26
Introduction	26
2.1 Memorable characters in the Dickensian world	27
2.2 Characterisation and text worlds	31
2.3 Induction and abstraction in the corpus stylistic circle	36
2.4 A text-driven approach to textual building blocks of fictional worlds	39
Conclusions	41
3 Starting with the Texts: Corpora, Clusters, and Lexical Bundles	42
Introduction	42
3.1 Texts and corpora	42

3.2	Defining and retrieving clusters—Initial examples	44
3.3	Approaches to clusters and functional interpretations	48
3.4	Lexical bundles or clusters for the study of Dickens's fiction?	51
3.4.1	Lexical bundles in fiction	51
3.4.2	Increasing the length of clusters	56
	Conclusions	60
4	Groups of Clusters for the Identification of Local Textual Functions	62
	Introduction	62
4.1	Quantifying, comparing, and classifying clusters	62
4.2	Key clusters and the distribution of clusters across texts	64
4.3	An operational definition of cluster groups	66
4.4	Distributions of cluster groups	67
4.4.1	Cluster categories in DCorp	68
4.4.2	Cluster categories in 19C	71
4.5	'Key groups' of clusters	73
	Conclusions	74
5	Character Speech	75
	Introduction	75
5.1	Functions of Speech clusters	76
5.1.1	Negotiating information	77
5.1.2	Turn-taking	78
5.1.3	Politeness formulae	80
5.1.4	First-person narration	81
5.1.5	Contextualising and highlighting functions	82
5.2	Phrasal impoliteness and confrontation	83
5.3	Functions in context: Pickwickian clashes and confrontation in <i>Oliver Twist</i>	90
5.4	Vague language and other examples of interpersonal meanings	95
	Conclusions	99
6	Body Language	100
	Introduction	100
6.1	Korte's (1997) framework for the analysis of body language	101

6.2	Body language and characterisation	103
6.3	From Body Part clusters to body language	105
6.4	The body language cline	110
6.4.1	The 'fireplace pose'	111
6.4.2	Gaze behaviour: <i>eyes</i>	114
6.4.3	Touch behaviour and authentication: <i>his hand upon his shoulder</i>	117
6.4.4	Empty body language?	118
6.4.5	Body language functioning as labels and thematic cues	119
	Conclusions	126
7	<i>As If</i> and the Narrator Comment	128
	Introduction	128
7.1	Clusters and fanciful meanings	128
7.2	Collocates and patterns of narrator comments	131
7.2.1	Action verbs, body part nouns, and settings	134
7.2.2	Manner	137
7.2.3	LOOK	141
7.2.4	SPEAK	143
7.2.5	Significant collocates—Extending the meaning groups	145
7.3	Lexically driven descriptions of body language	148
	Conclusions	151
8	Labels: Contextualising and Highlighting Functions	152
	Introduction	152
8.1	Groups of Labels and their functions	152
8.1.1	Reporting Speech Labels	153
8.1.2	Speech Labels	155
8.1.3	Body Language Labels	155
8.1.4	Attributes	155
8.1.5	Long clusters	156
8.1.6	Names of people and places	157
8.1.7	Labels and the highlighting and contextualising continuum	158
8.2	Labels in context: The example of Rigaud	159
	Conclusions	163

9 Conclusions and Outlook	164
9.1 Characters in the text-driven fictional world	164
9.2 The cluster and beyond	166
9.3 Suspensions	168
9.3.1 Suspensions and the <i>as if</i> cline	168
9.3.2 Concordancing suspensions	172
9.4 Towards a mixed methods approach	175
Appendices	
1 Texts in 19C	179
2 Texts in 36N	180
3 Cluster groups for shared key clusters (19C and 36N)	181
4 432 L clusters (occurring at least 5 times in one DCorp text) in alphabetical order	183
5 108 S clusters (occurring at least 5 times in one DCorp text) in alphabetical order	193
6 19 BP clusters (occurring at least 5 times in one DCorp text) in alphabetical order	196
7 14 AI clusters (occurring at least 5 times in one DCorp text) in alphabetical order	197
8 59 TP clusters (occurring at least 5 times in one DCorp text) in alphabetical order	198
9 84 d clusters (occurring at least 5 times in one DCorp text) in alphabetical order	200
<i>Notes</i>	203
<i>References</i>	209
<i>Index</i>	219

List of Tables, Figures, and Concordances

Tables

2.1 Areas of meanings in text-driven fictional worlds	40
3.1 Texts in DCorp (novels in <i>italics</i>)	43
3.2 Number of clusters in DC occurring at least 10 times	45
3.3 Top 20 of the two-, three-, and four-word clusters in DC	47
3.4 Five-word clusters in DC occurring a minimum of 10 times	47
3.5 Four-word lexical bundles in 37N	52
3.6 Four-word lexical bundles in 37Ns30	53
3.7 Top 10 five-word clusters in 37N and 37Ns30	57
3.8 Nine- to twelve-word clusters in DCorp and 19C	58
3.9 Ten-word clusters in DCorp	59
4.1 Key clusters for comparison of DCorp with 19C and 36N	64
4.2 Examples of key clusters in DCorp compared with 19C and 36N	65
4.3 Groups of clusters	67
4.4 Distribution of 56 key clusters across groups	67
4.5 Clusters in DCorp (grey = all five categories found, * = four categories found)	70
4.6 Clusters in 19C (grey = all five categories found, * = four categories found)	72
4.7 Key cluster groups for DCorp compared against 19C	73
5.1 Examples of confrontation clusters	89
5.2 <i>what do you mean by</i> in DCorp and 19C	91
5.3 <i>what do you want here</i> in DCorp	92
5.4 <i>what have you got to</i> in DCorp and 19C	92
6.1 Body Part clusters (freq. ≥ 5 and key clusters)	106
6.2 Occurrences of <i>Affery</i>	124
7.1 As If clusters	129
7.2 Top 25 left-collocates of <i>as if</i>	132

xii *List of Tables, Figures, and Concordances*

7.3 Meaning groups of left-collocates of <i>as if</i> (frequent collocates)	132
7.4 Meaning groups of left-collocates of <i>as if</i> (frequent and significant collocates)	146

Figures

1.1 The corpus stylistic circle	12
1.2 Units and levels of language description	14
1.3 The lexical item, its forms of realisation and levels of language description	16
1.4 Degrees of generalisation for the description of a lexico-grammatical core	18
2.1 Example of a text-world diagram	35
3.1 Five- to eight-word clusters in DCorp and 19C	58
4.1 Cluster categories in DCorp normalised per 100,000 words	69
4.2 Cluster categories in 19C normalised per 100,000 words	71
7.1 The <i>as if</i> cline of the externalisation of character	133
9.1 Screen shot of CLiC— <i>Mr. Snagsby</i> in suspensions	173
9.2 Working notes for chapter 1 of <i>Little Dorrit</i> (Stone 1987: 271)	176

Concordances

1.1 Examples of <i>true feelings</i> from a corpus of <i>Guardian</i> newspapers	15
5.1 <i>what have you got to say</i> in 19C	90
5.2 <i>not a bit of it</i> in DCorp	98
6.1 <i>with his back to the</i> in DCorp	112
6.2 <i>with his back to the</i> in 19C	113
6.3 <i>with her back to the</i> in DCorp	113
6.4 <i>with her back to the</i> in 19C	113
6.5 <i>(with) his eyes on the (ground)</i> in DCorp	115
6.6 <i>hand</i> with <i>shoulder</i> within a context of five words in NN	117
7.1 <i>air</i> meaning 'manner' or 'appearance' of a character in a five-word span to the left of <i>as if</i>	139

Introduction

[. . .] conventional phrases are a sort of fireworks, easily let off, and liable to take a great variety of shapes and colours not at all suggested by their original form.

—Dickens, *David Copperfield*, ch. 41

Dickens wrote *David Copperfield* long before corpus linguistics came into being, but he hits the nail on the head: from a corpus linguistic point of view, conventional phrases, or repeated patterns, constitute a most powerful aspect of the language. Corpus linguistics has provided the methodology to investigate repeated patterns in a new light, revealing facts about language that have largely remained hidden from human observation. This book proposes a corpus stylistic approach to the analysis of Dickens's fiction that seeks to combine corpus linguistics and literary analysis. The combination may in itself have enough potential for fireworks. Using computers to aid the analysis of literature does not seem an obvious choice. Nevertheless, there are several reasons why corpus stylistics is an attractive venture.

Corpus linguistics has had a major impact on the description of language. Central to corpus linguistic arguments is the focus on language in use. The availability of corpora, i.e., large collections of computer-readable texts, makes it possible to observe repeated patterns, and the patterns in turn serve as the basis for the description of repeatedly expressed meanings. The types of patterns retrievable with the help of corpora are often not consciously accessible by our linguistic intuitions. The most visible impact of the 'corpus revolution' has so far been seen in the field of lexicography. In the 21st century it seems to be almost mainstream that dictionaries are based on or supported by information from corpora. The implications of corpus work are even more far-reaching, and new models for the description of linguistic phenomena are emerging. However, with new models come new questions. One question is how useful corpus approaches are to the study of literary texts. After all, a poem, a novel, or a short story provide very limited data compared to corpora such as the British National Corpus containing about 100 million words—or even the web that is now used as a corpus. If corpus

2 *Corpus Stylistics and Dickens's Fiction*

linguistics aims to develop into a linguistic discipline in its own right, it needs to account for literature, too. Sinclair (2004) emphasises:

[l]iterature is a prime example of language in use; no systematic apparatus can claim to describe a language if it does not embrace the literature also; and not as a freakish development, but as a natural specialization of categories which are required in other parts of the descriptive system.

(Sinclair 2004: 51)

An important point here is that already the distinction into literary and non-literary language cannot be upheld (cf. Carter 2004). So, the question is not whether corpus approaches developed for non-literary language are also applicable to the literary variety of a language. What is at stake is finding adequate linguistic categories to work along the “cline of literariness in language” (Carter 2004: 69).

A corpus stylistic approach assumes that the linguistic analysis of a literary text provides useful insights complementing the literary appreciation of the work. However, the justification for the present book might have to start even earlier. Although no one would deny Dickens's linguistic skills, it seems that accounts of his language often focus on the identification of stylistic features in the form of overviews or inventories illustrated with isolated examples rather than on the linguistic analysis of fictional worlds. John (2001) highlights the gap in linguistic research and stresses the need for close analysis of Dickens's prose: “Dickens studies has failed to analyse adequately the melodramatic poetics informing Dickens's prose” (John 2001: 108). She points out that close reading is often assumed to “add nothing to currently central debates on ideology” and observes that “the historical association of criticism of the Victorian novel with all themes ‘big’ and social has not helped in this respect” (John 2001: 108ff.).

Corpus stylistics, as proposed in the present book, draws on corpus methodology but at the same time, it emphasises the link that literary stylistics provides to literary criticism. Thus, this book engages with concrete textual examples of Dickens's fiction and seeks to link the analysis of linguistic features to their functions in the creation of fictional worlds. Although the corpus stylistic approach that is proposed has wider applicability beyond one author, this book is very much focused on Dickens's fiction. There are at least three reasons for the choice to concentrate on works by Charles Dickens. First, there are some well-known features of Dickens's style. Brook (1970: 36) points out: “Dickens was never afraid of making excessive use of a way of writing that happened to appeal to him. [. . .] He was never afraid of repetition”. Second, Dickens's popularity shows that his way of writing did not only appeal to himself but also has been appealing to a wide audience right up to the present day. So there is plenty of evidence of the effects that Dickens's texts have had on their readers as well as literary criticism discussing these effects. Third, texts by