

A Corpus Linguistic Approach
to Literary Language and Characterization

Linguistic Approaches to Literature (LAL)

Linguistic Approaches to Literature (LAL) provides an international forum for researchers who believe that the application of linguistic methods leads to a deeper and more far-reaching understanding of many aspects of literature. The emphasis will be on pragmatic approaches intersecting with areas such as experimental psychology, psycholinguistics, computational linguistics, cognitive linguistics, stylistics, discourse analysis, sociolinguistics, rhetoric, and philosophy.

For an overview of all books published in this series, please see
<http://benjamins.com/catalog/lal>

Editors

Sonia Zyngier

Federal University of Rio de Janeiro

Joanna Gavins

University of Sheffield

Advisory Editorial Board

Douglas Biber

Northern Arizona University

Arthur C. Graesser

University of Memphis

Keith Oatley

University of Toronto

Marisa Bortolussi

University of Alberta

Frank Hakemulder

Utrecht University

Willie van Peer

University of München

Donald C. Freeman

University of Southern
California

Geoff M. Hall

University of Wales, Swansea

Yeshayahu Shen

Tel Aviv University

Richard Gerrig

Stony Brook University

David L. Hoover

New York University

Mick Short

Lancaster University

Raymond W. Gibbs, Jr.

University of California,
Santa Cruz

Don Kuiken

University of Alberta

Michael Toolan

University of Birmingham

Geoffrey N. Leech

Lancaster University

Reuven Tsur

Tel Aviv University

Rachel Giora

Tel Aviv University

Paisley Livingston

University of Copenhagen

Peter Verdonk

University of Amsterdam

Volume 18

A Corpus Linguistic Approach to Literary Language and Characterization
Virginia Woolf's *The Waves*
by Giuseppina Balossi

A Corpus Linguistic Approach to Literary Language and Characterization

Virginia Woolf's *The Waves*

Giuseppina Balossi

John Benjamins Publishing Company
Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Balossi, Giuseppina.

A corpus linguistic approach to literary language and characterization: Virginia Woolf's
The Waves / Giuseppina Balossi.

p. cm. (Linguistic Approaches to Literature, ISSN 1569-3112; v. 18)

Includes bibliographical references and index.

1. Language and languages--Style--Psychological aspects. 2. Cognitive grammar.
3. Discourse analysis, Literary. 4. Woolf, Virginia, 1882-1941. *The Waves*.
5. Psychological fiction, English--History and criticism. I. Title.

P301.5.P75B36 2014

823'.912--dc23

2014001873

ISBN 978 90 272 3407 0 (Hb; alk. paper)

ISBN 978 90 272 7042 9 (Eb)

© 2014 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Cui, nisi tibi?

*This book is dedicated to Antonio Nogara,
whose academic example, teaching and deep sensitivity
made me a different person*

Table of contents

Acknowledgements	XI
Foreword	XIII
List of conventions	XV
List of figures and tables	XVII
List of appendixes	XIX
List of concordances	XXI
CHAPTER 1	
Introduction	1
1.1 The issues at stake	1
1.2 The design of this book	2
CHAPTER 2	
Virginia Woolf's <i>The Waves</i>	5
2.1 Virginia Woolf and the modernist character	5
2.2 Introduction to <i>The Waves</i>	6
2.2.1 Conception	7
2.2.2 Structure	8
2.2.3 Method of characterization	9
2.2.4 Debate over characterization	11
2.2.5 Studies on lexical patterns	13
2.2.6 Psychoanalytic approaches to character	15
CHAPTER 3	
Literature review	19
3.1 The state of affairs in literary characterization	19
3.1.1 Characters as people and as textual constructs	19
3.1.2 Character typologies	22
3.2 Psychological concepts in the perception of personality	24
3.2.1 Prior knowledge and schema theory	25
3.2.2 History of schema theory	26

3.3	Stylistic approaches to characterization	28
3.3.1	Culpeper's model for characterization: Top-down and bottom-up	29
3.3.2	Cognitive metaphor, mind style and characterization	33
3.4	Studies of language and personality	35
3.5	Conclusion	39

CHAPTER 4

	Corpus approaches to the study of language and literature	41
4.1	Introduction	41
4.2	Corpus annotation	44
4.3	Techniques employed in corpus analyses	45
4.4	Corpus annotation of literary texts	49
4.4.1	Quantitative stylistic approaches to literary texts	51
4.4.2	Authorial style	52
4.4.3	Characterization	55
4.4.3.1	Characterization in <i>The Waves</i>	56

CHAPTER 5

	Methodology	59
5.1	Introduction	59
5.2	The narrative framework of <i>The Waves</i>	59
5.2.1	Synchronic and diachronic structure	60
5.2.2	<i>The Waves</i> as e-text: Text annotation and text division	64
5.3	Wmatrix: A software tool for corpus analysis and comparison	66
5.3.1	The CLAWS and USAS taggers and their tagsets	67
5.4	<i>The Waves</i> through Wmatrix	70
5.4.1	Adaptation and revision of POS and USAS tagsets	72
5.4.2	Data layout in Microsoft Excel	76
5.4.3	Post-editing of the soliloquy text	78
5.4.4	Token count of the soliloquy text	81

CHAPTER 6

	Character differentiation through word-classes	83
6.1	Introduction	83
6.2	Word-class and characters' language	83
6.2.1	Word-classes in the whole soliloquy text and in each character	85
6.2.2	Data comparison and treatment of the LL statistics	89
6.2.3	Characters' word-classes in each phase of life: Significant differences relative to each character	93
a.	Susan	96
b.	Jinny	96

c. Rhoda	97
d. Louis	97
e. Neville	98
f. Bernard	98
6.2.4 Bernard's word-classes in the final soliloquy	99
6.3 Summary	102
CHAPTER 7	
Character differentiation: Semantic fields	107
7.1 Introduction	107
7.2 Data comparison and treatment of the LL statistics	108
7.3 The characters' schema of gender dichotomy	112
7.4 The female characters	114
7.4.1 Susan	115
a. The Natural World	116
b. The House	120
c. Home-making activities	121
d. Family and Motherhood	122
e. Time	123
f. The Senses	125
7.4.2 Jinny	127
a. The Body	127
b. Colours	130
c. Movement	131
7.4.3 Rhoda	133
a. The Natural World	134
b. Movement	136
c. Emotions	139
d. Shapes	140
e. Colours	142
7.5 The female characters' semantic fields and personality traits	146
7.6 The male characters	150
7.6.1 Louis	150
a. Geographical names	151
b. People and Relationships	152
c. The Natural World	158
d. History and Literature	159
7.6.2 Neville	161
a. Literature and Philosophy	161
b. Academic Interests	165

7.6.3	Bernard	166
a.	Cognition and Abstraction	167
b.	Intellectual Interests	172
7.7	The male characters' semantic fields and personality traits	175
7.8	Bernard's final summing-up and his dramatic role	178
7.9	Conclusion: The characters' semantic fields and their personality traits	182
7.10	Percival: The other-presented character	183
CHAPTER 8		
	Conclusion	185
8.1	Achievements	185
8.2	Characterization in this study	186
8.3	Reading through the computer	187
8.4	Authorial style <i>versus</i> character voice	188
8.5	Further directions	190
	References	193
	Websites	211
	Appendixes	213
	Author index	273
	Subject index	275

Acknowledgements

I should like to express my gratitude to a number of people who have contributed to the ‘birth’ of this book. Special thanks are due to Olivia Fialho, whom I met at the annual conference of the Poetics and Linguistics Association (PALA) in Genoa in 2010. It was through her that my research reached Sonia Zyngier. And it is thanks to Sonia’s support and patience that I had the time to bring this work to its completion. I am especially grateful to Elena Semino, who supervised both my MA and Ph.D. theses on characterization at Lancaster University. I would also like to express my gratitude to Jonathan Culpeper, who supervised me while Elena was on sabbatical. Without their help and encouragement, I could not have gone this far.

I would also like to acknowledge the help of several academic colleagues I met at Lancaster University who have now taken different academic paths. Nicholas Smith and Amy Wang accompanied me throughout, Nick helping with the statistics and Amy giving much valuable advice. Michael Oakes also gave me kind, helpful support with the statistical part of the work, always ready to settle my doubts. Izumi Tanaka very kindly helped with the tagging of the text and Paul Rayson with detailed information about Wmatrix whenever I needed it.

I am also grateful to friends from other walks of life, especially Mike Cardwell, who, besides giving moral support, also spent a huge amount of time straightening out the layout of the data and providing practical advice on how to deal with the very large datasets. I am also grateful to Adrian Paul and Christina Chow, Andrew Wilson, Emily Bianchi-Bazzi, Giangio Gilardi, Massimo Vaghi, and my student Loris Molinari, who have either helped with some parts of the book or provided general and specific comments.

Last but not least, I am deeply grateful to John Heywood, who first proof-read the Ph.D. thesis from which this book developed and then supported me throughout the writing process from the first draft to the very last. Not only did he read and re-read it through carefully, but he also made several important editorial suggestions, especially regarding metaphor analysis. He commented thoroughly and patiently on the style and presentation of each chapter, and helped with the bibliography and index. This long collaboration has also contributed to our deep affective bond. It has been an extraordinary experience to work with him.

This book began while my father was still with us and I am aware of how proud he would have been of its completion, as my mother is today. I would also like to thank the assistance I had from my sisters, Mari and Tina, who did more than their fair share during my withdrawal from family life.

Finally, I owe my deepest gratitude to my husband Mario and my daughter Fabiola, who have kept me on track throughout the project and had often had to do without me. To them I offer the seeds I have sown.

Foreword

“Language use is one of the most mysterious products of the human mind, and literature is probably the most enigmatic form of language use” (van Peer 2011: 1). Before the arrival of structuralism, the study of literature tended to be based on free, intuitive interpretation. Today, with the advent of the computer, more systematic studies can be carried out. Will they offer possibilities for solving the enigma of language in literature?

While computer-assisted methods of enquiry have been familiar for some time now across a range of disciplines, such as psychology and other social sciences (Murphy 2010; Pennebaker 2011), the field of literary criticism has resisted the lure of the computer (Gottschall 2008). Corns (1991: 129) observed that computer-aided literary analyses were not being paid sufficient attention and unfairly associated with poor achievements, whereas stylometric studies (Burrows 1987) formed a distinct strand kept quite separate from the main literary critical tradition. Louw (1993: 152–176; 1997: 240–251) was one of the first scholars to appreciate the potential value and effectiveness of corpus linguistics in the analysis of literary texts. Now computer-aided studies have been accepted as a major research paradigm (Stockwell 2008: 351–363) that can contribute a great deal to developing fresh critical insights into the stylistic qualities of texts, including literary ones (Deignan 2005; Zyngier 2008: 169–190; Lüdeling and Kytö 2008; 2009; Starcke 2010; Viana *et al.* 2011).

While we may all agree that the computer will never be able to understand and appreciate a literary text in the ways the human mind does (Graesser *et al.* 2011: 24–33), we must acknowledge that corpus methodologies open up, in Leech’s words (1992: 106), “a new way of thinking about language.” Using the computer leads to discoveries that cannot be seen with the naked eye. A computer-aided approach can also offer new starting points for investigation. It has proven to be a strong research method anchored firmly in theoretical frameworks which can be combined with other qualitative methodologies (Leech and Fligelstone 1992: 115–140; Biber *et al.* 1998; Biber 2011: 15–23; Mahlberg and McIntyre 2011: 204–227; McIntyre 2012: 409–415).

This book attempts to demonstrate how a computer-aided approach can guide us through the investigation of language and characterization in Virginia Woolf’s *The Waves* (1931). The focus is on whether characters can be differentiated

through examining the word-classes and semantic categories associated with them. Although these corpus-based findings may sometimes be problematic, it is argued that they produce new and detailed insights into both the diachronic development of the characters and the novel as a whole. While the investigation of the word classes conducted to assess the presence of individuality in the characters' voices or idiolects (Stewart 2003: 129–138) is purely quantitative, the semantic analysis is more finely articulated. By applying Culpeper's (2001) cognitive model of characterization, the semantic fields distinguishing each character to a statistically significant extent are also subjected to qualitative analysis, allowing inferences to be drawn about probable differences in personality traits and idiosyncratic mind styles (Fowler 1977: 103–113; Semino 2007: 153–203).

The book is conceived in such a way as to bring both theory and practice together. Background information is provided on the theoretical aspects of the debate over characterization in *The Waves*, followed by a broad survey of studies of characterization from their outset to the present state of the art, and of proposals for viable ways to study fictional people's language and their personalities (Eder *et al.* 2010). The study also provides and puts into practice a detailed methodology for the analysis of character. The corpus-based approach, as suggested here, is also a methodology applicable to the contrastive investigation of language, and of language and personality, in different types of discourses encompassing both fictional *personae* and real people.

List of conventions

ACAMRIT	Automatic Content Analysis of Market Research Interview Transcripts
ACASD	Automatic Content Analysis of Spoken Discourse
BNC	British National Corpus
CAMET	Computer Archive of Modern English Texts
CLAWS	Constituent Likelihood Automatic Word-tagging System
CMT	Cognitive Metaphor Theory
LL	log-likelihood
LOB	Lancaster/Oslo-Bergen Corpus
MDA	multi-dimensional analysis
MVA	multi-variate analysis
MWE	multi-word expression
OTA	Oxford Text Archive
POS	part-of-speech
REVERE	REVerse Engineering of Requirements project
SEMTAG	Semantic Tagger
SGML	Standard Generalized Mark-up Language
SP	semantic prosody
UCREL	[Lancaster] University Centre for Computer Corpus Research on Language
USAS	UCREL Semantic Analysis System
WD	<i>The Diaries of Virginia Woolf</i>

List of figures and tables

Tables

Table 5.1	The characters' life stages and major events	62
Table 5.2	Text files used in the analysis of the soliloquy part of <i>The Waves</i>	65
Table 5.3	The top semantic levels in the USAS system	69
Table 5.4	The semantic subfields in USAS level 'B: The body and the individual'	70
Table 5.5	The top semantic levels in the modified USAS system	73
Table 5.6	The data layout taken from Susan's speaking section in ALLCHRS POS and USAS dataset	77
Table 5.7	Token size in the soliloquy part of <i>The Waves</i>	82
Table 6.1	Word-class frequency distribution in the soliloquy part of <i>The Waves</i>	85
Table 6.2	Comparison between word-class frequencies in the soliloquy part of each character's spoken text	87
Table 6.3	Sample output of the statistically significant LL values in Susan's word-classes compared with Jinny's word-classes in Speaking Section 2	91
Table 6.4	Sample output of the statistically significant LL values in Susan's word-classes compared with Jinny's word-classes in each phase of life	92
Table 6.5	Bernard's significant word-classes in the final summing-up	99
Table 7.1	Sample analysis of the LL statistics comparing Susan's semantic fields with the other characters' semantic fields	110
Table 7.2	Sample analysis of the frequency distributions of Susan's most overused semantic fields in each phase of life	111
Table 7.3	Susan's most overused semantic fields relative to the other characters	115
Table 7.4	Jinny's most overused semantic fields relative to the other characters	127
Table 7.5	Rhoda's most overused semantic fields relative to the other characters	134
Table 7.6	Louis' most overused semantic fields relative to the other characters	151
Table 7.7	Frequencies of occurrence of names in Louis' semantic field tag Z1.2: Minor characters' personal names	156
Table 7.8	Neville's most overused semantic fields relative to the other characters	161
Table 7.9	Neville – frequencies of occurrence of names in the semantic field tag Z1: Personal names	162
Table 7.10	Bernard's most overused semantic fields relative to the other characters	167
Table 7.11	Bernard's most significant semantic fields in the summing-up	179
Table 7.12	Frequency of occurrence of the most frequent items in L1: Life and living things in Bernard's Section 18 compared to Sections 2–16	180

Figures

Figure 6.1 Each character's corpus size in all speaking sections 86

Note. All Figures and Tables that are too large to fit into the body of the book are listed as Appendixes.

List of appendixes

- Appendix A. Log-likelihood calculator
- Appendix B. UCREL semantic tagset and UCREL modified semantic tagset
- Appendix C. ALLCHRS POS and USAS
- Appendix D. Each character's statistically significant LL values for the word-classes in each speaking section
- Appendix E. Number of statistically significant differences in each character's life stage
- Appendix F. Susan's significant differences relative to each character
- Appendix G. Jinny's significant differences relative to each character
- Appendix H. Rhoda's significant differences relative to each character
- Appendix I. Louis' significant differences relative to each character
- Appendix J. Neville's significant differences relative to each character
- Appendix K. Bernard's significant differences relative to each character
- Appendix L. Each character's semantic fields in all speaking sections
- Appendix M. Each character's overused semantic fields
- Appendix N. Each character's frequency distribution for the most overused semantic fields in each phase of life

List of concordances

Concordance 7.1	Semantic field tag S4: “father”	123
Concordance 7.2	Semantic field tag T1.3: “day”	124
Concordance 7.3	Semantic field tag X3.4: “see”	126
Concordance 7.4	Semantic field tag B2: “dress”	128
Concordance 7.5	Semantic field tag B1: “body”	129
Concordance 7.6	Semantic field tag L3: “flowers”	135
Concordance 7.7	Semantic field tag L2: “tiger”	138
Concordance 7.8	Semantic field tag O4.2: “dark/darkness”	143
Concordance 7.9	Semantic field tag Z1.1: “Bernard”	153
Concordance 7.10	Semantic field tag Z1.1: “Rhoda”	155
Concordance 7.11	Semantic field tag X8: “try/tried”	157
Concordance 7.12	Semantic field tag X2: “feel”	168
Concordance 7.13	Semantic field tag A6: “another/other”	169
Concordance 7.14	Semantic field tag Q2: “story/stories”	173
Concordance 7.15	Semantic field tag Q1.2: “phrases/phrase”	174

Introduction

1.1 The issues at stake

Considered one of the most astonishing narrative achievements of the twentieth century Modernist novel and Virginia Woolf's masterpiece, *The Waves* (1931) is perhaps her most controversial and difficult work due to its unique and original narrative structure and mode of character presentation. Indeed, the characters in this novel use a highly symbolic and artificial style of language that is not found in ordinary communication. Moreover, they all appear to speak in this similar style throughout the story. All of these features combine to set this work apart not only from the Victorian novel, but also from other experimental narrative writings of the twentieth century (Gorsky 1972: 449–466; Leaska 1984: 102–103; Flint 1994: 219–247; Hite, in Woolf 2006: xxxvi–lxvii). All this has played a role in stimulating considerable debate over the method of characterization in this novel.

In *The Waves*, the lives of six characters, three females and three males, are presented by means of soliloquies delivered by each at different stages of their lives from childhood to old age. An additional character, Percival, is also present, though he never speaks and is entirely other-presented. Each stage of life is framed by an interlude consisting of a brief stretch of lyrical description written in third-person narration (Graham 1976: 14). In each interlude, a part of the day is described, moving from sunrise to sunset in sequence, corresponding symbolically to the life stage depicted in the soliloquy section following. Within each soliloquy section, the characters' individual soliloquies are clearly demarcated by reporting clauses (e.g. "said Susan," "said Jinny," "said Rhoda," etc.). Although these make clear who is speaking, several critics have claimed that in the six characters' soliloquies (1) their language is identical in style, and that (2) this style remains stable throughout the narrative. They go on to claim that, for these reasons, we cannot distinguish whether (3) each speaker can be attributed an individual identity, or whether (4) each displays any signs of personal development from their childhood to their old age (see, for example, Guiguet 1965: 280–302; Graham 1970: 28–44; Minow-Pinkney 1987: 152–186; Caughie 1991: 62–90).

In the light of these claims, the aim of this book is to investigate whether the language each of the six speakers uses distinguishes each of them from each other. This broad issue is tested by means of a series of statistical comparisons of each character's word-classes and semantic fields. As will be explained in the description

of the design of this book, the approach adopted is computer-aided. Specifically, in order to investigate the various aspects of language, Wmatrix tools for word-class and semantic analysis tagging are used.

The word-class analysis is quantitatively based, since the characters' use of word-classes is measured statistically according to the most frequently used word-classes in each phase of their lives. The investigation of the semantic properties of their language is more complex and finely articulated. The first aim here is to establish the presence of any statistically significant differences in the characters' use of different semantic fields. The point of departure is similar to that applied in the word-class analysis, in that I first identify the most key fields distinguishing each speaker from the others. However, it is also shown how these findings can be used qualitatively to suggest how we may attribute different personality traits to them.

1.2 The design of this book

Chapter 1 pins down the state of the art in the literary critical debate on characterization in Virginia Woolf's *The Waves*. It introduces the issues raised by characterization in this work, the quantitative approach applied to it and provides an outline of this book. Chapter 2 presents Virginia Woolf as a modernist writer and her involvement with developing a new way of writing fiction, especially concerned with the approach to characterization in fiction. The chapter looks at the difficult genesis of *The Waves* and at its literary critical reception, both positive and negative, from the time of its publication until the present day. Then, it moves on to describe its narrative structure and Woolf's method of constructing characters in this novel. In particular, it discusses the opposing arguments raised by literary critics with regard to the ontological status of its characters, their differentiation and development. The features of language and some of the studies carried out in this area are also discussed. The chapter concludes with some observations on the psychoanalytic approaches that have supported part of the literary critical, and especially the feminist, approaches to the interpretation of Woolf's works, and also of *The Waves*. Chapter 3 provides an overview of the state of affairs in theories of literary characterization and describes the different theories put forward in regard to whether literary characters should be viewed as textual products, as the great majority of formalist, structuralist and semiotic scholars have argued, or whether literary characters can be discussed as having human-like aspects. Next, the discussion presents the range of competing character typologies available. Starting from the pioneering distinction made by E. M. Forster between "flat" and "round" types, it looks in succession at how subsequent critics have criticized this categorization and at how they have augmented his character types and integrated them into more complex typologies. The chapter then goes on to consider the potential of psychology and social

psychology to contribute to the characterization debate. It focusses on the definition of personality and outlines how it can be inferred *via* social schemata. Such mental structures not only help in forming impressions about the people we meet in everyday life, but are also relevant to the perception of fictional *personae*. This leads onto a discussion of some notions from psychology regarding how background knowledge is said to be organized in the mind, paying attention to schema theory, its interdisciplinary origins and theoretical development. Within cognitive stylistics, character comprehension is based on the idea that we understand people's personalities through the activation and interaction of top-down processes of inferencing (based on prior knowledge organized in social schemata or stereotypes) and bottom-up processes (derived from textual cues). After that the chapter presents the cognitive approach to characterization, in particular as articulated by Culpeper, whereby character is viewed as a construct that arises from the interaction between the reader and the text. The remaining part of the chapter considers studies in language and personality within social psychology that are mainly conducted through bottom-up approaches and through computer-aided techniques. Scholars within this strand acknowledge that investigations in social cognition have contributed a great deal of knowledge on how our minds encode, store and retrieve information and on how social schemata are first formed and subsequently used in the perception of social groups. However, they integrate this traditional view with a new attention to discourse by claiming that social schemata or stereotypes are inherent in the language we use to describe people, rather than in mental pictures in our mind. They regard words as a powerful tool through which we share information about individuals by describing their traits and thus pass down our beliefs and stereotypes about them to others. Chapter 4 illustrates how and why a computer-aided approach represents an effective methodology for studying characters' language and ascertaining their linguistic differentiation. The chapter provides the definition of the term 'corpus' and points out that it has a rather specific and atypical meaning in the present study, referring, namely, to either the whole text of *The Waves*, or, more often, to the separate sections uttered by the six characters. Although this is not a prototypical use of the term 'corpus,' which normally implies a carefully balanced and controlled selection of samples of multiple texts, the practice is not unprecedented. In fact, it has been used by scholars who, as here, have applied corpus methodologies to the stylistic analysis of a single literary text. The chapter proceeds with the major characteristics of the corpus linguistic approach to the study of language in general, the types of corpus annotations and the quantitative and qualitative techniques used to study corpora. It continues with a description of corpus-based approaches to the study of literary texts and points out the differences that are specific to their application to literary texts. In succession, it takes into account the quantitative studies applied in the area of authorship attribution, characterization in general and in *The Waves*. Chapter 5 deals with the methodology

involved in the preparation and annotation of the text of *The Waves* before it could be put into Wmatrix, and with the modifications to and arrangement of the data obtained after the processing phase. Chapter 6 presents the macro-analysis of the word-classes in the soliloquy part. It firstly compares the frequency distributions of word-classes in each character's spoken text with that of all the soliloquies taken as a whole. This is carried out in order to investigate to what extent the word-class frequency distribution of each character varies in relation to that of the entire soliloquy part, and to see whether there is variation within each character's soliloquies across their different stages of life. It then moves on to the statistical comparison of each character's word-classes in each life stage. The chapter concludes with a summary of the findings that have emerged from the statistical comparison of the word-classes and discusses some issues arising from the analysis. Chapter 7 investigates the statistically significant differences in the characters' use of different semantic fields. Here, the central question is whether, in spite of the alleged homogeneity of speech style displayed, the characters may be attributed individual traits and development across their lives. Greater emphasis is laid on the textual cues that may influence bottom-up processing. More specifically, the chapter reflects on how the semantic patterns revealed may lead to the attribution of different personality traits and show how these change along life. The chapter also takes into account the schema for gender dichotomy derived from the social categories evident in the cultural and historical context of the first half of the twentieth century, which some literary critics have viewed as underlying this novel. The top-down inferences are mainly derived from this information. The metaphorical patterns, which emerge from the detailed analysis, are interpreted according to cognitive metaphor theory (CMT) and, wherever appropriate, to reveal the idiosyncratic world-views of each of the six speakers. Finally, the chapter summarizes the most idiosyncratic lexical features that differentiate each character in their use of language, and the individual traits inferred from them in the course of the diachronic analysis. It also relates them to that strand of the literary criticism on characterization in *The Waves* that, through non-quantitative approaches, has held that the characters do indeed display differentiation in their use of language and variations in their personality traits across life. A summary of the results obtained is finally illustrated pointing out that the statistical analysis of the top significant semantic fields provides some important evidence for the potential attribution of distinct personalities. Moreover, it highlights how the diachronic analysis of such findings can account for the changes in our impression formation, whereby each character is perceived, to a greater or lesser extent, as dynamic, and not as static. Chapter 8 re-considers a series of issues encountered in the application of a computer-aided approach to this novel. The book concludes with some suggestions for further directions that could be taken in the study of language and character through the computer.

Virginia Woolf's *The Waves*

2.1 Virginia Woolf and the modernist character

Virginia Woolf's contribution to Modernism (Rantavaara 1953: 150–160; Kumar 1963: 64–102) is not only confined to her activity as a novelist but also extends to the theory of the novel.¹ Proof of this is shown in her many essays, autobiographical reflections and diaries (Bell 1977–1984). In particular, she wrote much towards the shaping of a new kind of literary character. In her famous essay *Mr Bennett and Mrs Brown* ([1924] 1984: 192–212; see also Abbott 1993: 393–405), she held that the basis of good fiction lies in the quality of its presentation of character. She attacked the materialist Edwardians such as Bennett, Wells and Galsworthy, whose main concern was to portray their characters from the outside, describing their actions rather than their thoughts, in a series of well-defined chronological events, all carefully governed by the presence of the authorial voice. According to Woolf, the task of a good novelist was, instead, that of turning inwards to explore the individual's complex consciousness. This meant that she had to find a new method of writing fiction, which in her essay *The Narrow Bridge of Art* (Woolf [1958] 1975: 18–19), she suggested should

[...] resemble poetry in this that it will give not only or mainly people's relations to each other and their activities together, as the novel has hitherto done, but it will give the relation of the mind to general ideas and its soliloquy in solitude. For under the dominion of the novel we have scrutinized one part of the mind closely and left another unexplored. We have come to forget that a large and important part of life consists of our emotions towards such things as roses and nightingales, the dawn, the sunset, life, death and fate [...]. The psychological novelist has been too prone to limit psychology to the psychology of personal intercourse [...]. We long for some more impersonal relationship. We long for ideas, for dreams, for imaginations, for poetry.

1. For an overview and criticism of Woolf's novels, see Peel 1933: 78–96; Blackstone 1949; Daiches 1960; Bennett 1964: 91–111; Guiguet 1965; Briggs 1994; Lee 1977; Marcus 1997. The collected articles in Briggs and the book by Marcus describe Woolf's works mainly from a feminist perspective. Especially relevant, for the study of *The Waves* and characterization, are the essays *Mr Bennett and Mrs Brown* ([1924] 1984: 192–212) and *Modern Fiction* ([1925] 2008). For a complete and exhaustive bibliography of Virginia Woolf's works, see Kirkpatrick and Clarke (1997). For a complete autobiography of the author, see Bell (1973) *Virginia Woolf: A Biography*.

Despite the fact that Woolf worked hard in her writing to realize this idea, critics have generally remarked on her “inability to create characters” (Guiguet 1971:40), often defining them as caricatures or types (Bennett 1964: 19–41). E. M. Forster (in Bennett 1970: 17) observed how “she could seldom so portray a character that it is remembered afterwards on its own account.” Muller (1937:4) called them “disembodied” and attributed such a flaw to her excessive stress on interiority. In contrast, T. S. Eliot (1941: 313–316) praised the great gift that she had in portraying their inner minds.

2.2 Introduction to *The Waves*

The Waves represents Woolf’s concern to develop a new form of writing based on the centrality of character and the mental processes associated with it. Published in September 1931, this novel is today often regarded as her masterpiece, as well as one of the most remarkable narrative achievements of the twentieth century Modernist fiction. Yet, on its publication, Woolf was not spared negative criticism.

Most people are going to find *The Waves* extremely difficult reading – all people, in fact, excepting those who are prepared to accept the author’s highly artificial trick in writing it for the sake of the poetic images she invokes [...] it is hard to see why Mrs Woolf chose so odd a manner to convey what she had to say. [The characters] are simply six Mrs Woolfs, they are not more than attenuated shadows – brilliant, many-sided, tricky, but still shadows – of the real people the reader has a right to expect [...].

(Unsigned review, *San Francisco Chronicle*, 6 December 1931)

However, there were reviewers and friends who were able to capture its true value.² Amongst them, we find Nicolson (*Action*, 8 October 1931:6), who wrote,

[t]here is a note in this book which has never yet been heard in European literature [...]. Her whole intention is to depict the fluidity of human experience, the insistent interest of the inconsequent, the half-realised, the half-articulate, the unfinished and the unfinishable [...]. The book is difficult. Yet it is superb.

More recent critics and scholars have also stressed its complexity and difficulty; some of them have praised its high stature, others have considered it as Woolf’s most experimental work and have remarked on its formal innovation. Van Buren Kelley (1973: 144) notes that “[o]f the eight major novels that Woolf wrote, *The Waves* is by far the most intricate, both structurally and thematically.” Boyd

2. Excerpts of these reviews and others can be accessed at <http://www.uah.edu/woolf/waves.html>. Goldman (1997:21–28) provides a detailed overview of the contemporary reviews and comments to the publication of *The Waves*.

(1983: 93) views it as Woolf's "most radical break with realism," Payne (1969: 209) as a work of "puzzling complexity," Bishop (1991: 115) asserts that in it Woolf "goes to her furthest extreme in exploring the limits of language." McNichol (1990: 140) attributes the difficulty of understanding this work to the fact that it "is a mystical poetical work of verbal complexity on a ground base of simplicity and profundity." Daiches (1960: 215) sounds a negative note when he writes that it "displays more conscious virtuosity than any other of Woolf's novels." In contrast, Treloar (1988: 2.1)³ observes that it is Woolf's "*magnum opus*" in which "the synthesis of her views on personality, unity in diversity, the nature of humanity" is "presented in a vehicle quite unlike anything else in the English language."

2.2.1 Conception

Virginia Woolf began to think about *The Waves* as early as 1926 (see also Hussey 1995: 351–352), when in her diary (hereafter WD), she noted down that she wanted to write "some semi mystic very profound life of a woman, which shall all be told on one occasion; and time shall be utterly obliterated" (WD, 23 November 1926: 118). This idea kept hovering in her mind throughout 1927 and 1928, while she was writing *Orlando* (1928) and *A Room of One's Own* (1929) (a detailed account of the writing process at the basis of this work is given by Graham (1976), who edited the two holograph drafts). In the two diary entries below, we find references to it, which she originally referred to as "*The Moths*" (WD, 18 June 1927: 139).

Slowly ideas began trickling in; & then suddenly I rhapsodised [...] & told over the story of the Moths, which I think I will write very quickly, perhaps in between chapters of that long impending book on fiction. Now the moths will I think fill out the skeleton which I dashed in here: the play-poem idea: the idea of some continuous stream, not solely of human thought, but of the ship, the night &c, all flowing together: intersected by the arrival of the bright moths.

(WD, 18 June 1927: 139)

Yes, but *The Moths*? That was to be an abstract mystical eyeless book: a playpoem.

(WD, 7 November 1928: 203)

What Woolf wrote reflects her concern with the form ("a playpoem"), content ("abstract") and the narrative technique ("eyeless," *i.e.* without an 'I' narrator) that her book ("*The Moths*") should have. It was not until 1929 that she began seriously on what she now also started to call "*The Waves*" (WD, 11 October 1929: 259; 23 October 1929: 262). The original idea of "a mystical eyeless" work was turned

3. References to Treloar (1988), which was accessed as an e-text, are in the form of chapter, section and sub-section numbers.

into that of “[a] mind thinking,” a “She.” Moreover, the narrative structure that the book was to take in the final drafts started to become apparent as two “different currents,” which were to be called “interludes” and “dramatic soliloquies” (WD, 28 May 1929: 312; see 2.2.2 for a detailed account of the structure of the novel).

A few months later, the feminine identity of the narrator was excluded from the original design and Woolf wrote: “[w]ho thinks it? And am I outside the thinker? One wants some device which is not a trick” (WD, 25 September 1929: 257). The questions of who tells the story and how to tell it (“some device”) remained unsettled until the very last draft when Woolf resolved to present the life of six people in “a series of dramatic soliloquies [...] running homogeneously in and out, in the rhythm of the waves” (WD, 20 August 1930: 312).

2.2.2 Structure

The final version of the novel consisted of two interwoven parallel narrations: the interludes and the dramatic soliloquies. The interlude sections are nine in total and written in italics. Until the very last draft, the interludes still followed the soliloquies (Heine 1972: 60–80). They describe the progression of a day from dawn to sunset through an external observer or impersonal narrator. Each interlude is followed by a soliloquy section in normal type; here six individuals are presented (Susan, Jinny, Rhoda, Louis, Neville and Bernard), from childhood to old age, through consecutive utterances in the form of direct speech. Each section of direct speech is rigorously indicated by reporting clauses of the form ‘said X,’ as for instance “I see a slab of pale yellow,” said Susan, [...] “I hear a sound,” said Rhoda” (Woolf 1931: 7; for the edition adopted in the present study, see 5.2.2). A seventh character, Percival, is also present, but he is only ever indirectly presented through the other speakers later on in the story and dies before the end of it. The last soliloquy departs from the preceding ones in that it is abnormally long and spoken only by Bernard. His long final speech is followed by the final brief appearance of interlude text in the form of a single italicized sentence – “*The waves broke on the shore.*” (id.: 212) – which brings the novel to an end.

The striking departure from Woolf’s preceding novels is evident. For instance, in *Mrs Dalloway* (1925) and *To the Lighthouse* (1927), although Woolf had used interior monologue with its flashback technique to show individual consciousness, she had still relied on the presence of an omniscient narrator with the authority to step into the narration in order to explain what is going on. In the soliloquy part, the presence of the narrator is reduced to ‘said X.’ The fact that Woolf conceived it as a “playpoem” (WD, 7 November 1928: 203) hints at her intention to write a work sharing the qualities of poetry and drama. The first feature is manifest in the