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

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)

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

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 proofread 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	*The Diaries of Virginia Woolf*

List of figures and tables

Tables

Figures

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

List of appendixes