Language, Cohesion and Form

Margaret Masterman was a pioneer in the field of computational linguistics. Working in the earliest days of language processing by computer, she believed that meaning, not grammar, was the key to understanding languages, and that machines could determine the meaning of sentences. She was able, even on simple machines, to undertake sophisticated experiments in machine translation, and carried out important work on the use of semantic codings and thesauri to determine the meaning structure of text. This volume brings together Masterman's groundbreaking papers for the first time. Through his insightful commentaries, Yorick Wilks argues that Masterman came close to developing a computational theory of language meaning based on the ideas of Wittgenstein, and shows the importance of her work in the philosophy of science and the nature of iconic languages. Of key interest in computational linguistics and artificial intelligence, this book will remind scholars of Masterman's significant contribution to the field.

Permission to publish Margaret Masterman's work was granted to Yorick Wilks by the Cambridge Language Research Unit.

YORICK WILKS is Professor in the Department of Computer Science, University of Sheffield, and Director of ILASH, the Institute of Language, Speech and Hearing. A leading scholar in the field of computational linguistics, he has published numerous articles and six books in the area of artificial intelligence, the most recent being *Electric Words: Dictionaries, Computers and Meanings* (with Brian Slator and Louise Guthrie, 1996). He designed CONVERSE, the dialogue system that won the Loebner prize in New York in 1998.

Studies in Natural Language Processing

Series Editors:
Steven Bird, University of Melbourne
Branimir Boguraev, IBM, T. J. Watson Research Center

This series offers widely accessible accounts of the state-of-the-art in natural language processing (NLP). Established on the foundations of formal language theory and statistical learning, NLP is burgeoning with the widespread use of large annotated corpora, rich models of linguistic structure, and rigorous evaluation methods. New multilingual and multimodal language technologies have been stimulated by the growth of the web and pervasive computing devices. The series strikes a balance between statistical versus symbolic methods, deep versus shallow processing, rationalism versus empiricism, and fundamental science versus engineering. Each volume sheds light on these pervasive themes, delving into theoretical foundations and current applications. The series is aimed at a broad audience who are directly or indirectly involved in natural language processing, from fields including corpus linguistics, psycholinguistics, information retrieval, machine learning, spoken language, human–computer interaction, robotics, language learning, ontologies and databases.

Also in the series
Douglas E. Appelt, *Planning English Sentences*
Madeleine Bates and Ralph M. Weischedel (eds.), *Challenges in Natural Language Processing*
Steven Bird, *Computational Phonology*
Peter Bosch and Rob van der Sandt, *Focus*
Pierette Bouillon and Federica Busa (eds.), *Inheritance, Defaults and the Lexicon*
Ronald Cole, Joseph Mariani, Hans Uszkoreit, Giovanni Varile, Annie Zaenen, Antonio Zampolli and Victor Zue (eds.), *Survey of the State of the Art in Human Language Technology*
David R. Dowty, Lauri Karttunen and Arnold M. Zwicky (eds.), *Natural Language Parsing*
Ralph Grishman, *Computational Linguistics*
Graeme Hirst, *Semantic Interpretation and the Resolution of Ambiguity*
András Kornai, *Extended Finite State Models of Language*
Kathleen R. McKeown, *Text Generation*
Martha Stone Palmer, *Semantic Processing for Finite Domains*
Terry Patten, *Systemic Text Generation as Problem Solving*
Ehud Reiter and Robert Dale, *Building Natural Language Generation Systems*
Manny Rayner, David Carter, Pierette Bouillon, Vassilis Digalakis and Mats Wirén (eds.), *The Spoken Language Translator*
Michael Rosner and Roderick Johnson (eds.), *Computational Linguistics and Formal Semantics*

Richard Sproat, *A Computational Theory of Writing Systems*
George Anton Kiraz, *Computational Nonlinear Morphology*
Nicholas Asher and Alex Lascarides, *Logics of Conversation*
Walter Daelemans and Antal van den Bosch, *Memory-based Language Processing*
Margaret Masterman (edited by Yorick Wilks), *Language, Cohesion and Form*

Language, Cohesion and Form

Margaret Masterman
Edited, with an introduction and commentaries, by

Yorick Wilks

Department of Computer Science
University of Sheffield

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521454896

First published 2005

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this book is available from the British Library

ISBN-13 978-0-521-45489-6 hardback
ISBN-10 0-521-45489-1 hardback

Contents

*Starred chapters have following commentaries by the editor (and by Karen Spärck Jones for chapter 6).

Preface

This book is a posthumous tribute to Margaret Masterman and the influence of her ideas and life on the development of the processing of language by computers, a part of what would now be called artificial intelligence. During her lifetime she did not publish a book, and this volume is intended to remedy that by reprinting some of her most influential papers, many of which never went beyond research memoranda from the Cambridge Language Research Unit (CLRU), which she founded and which became a major centre in that field. However, the style in which she wrote, and the originality of the structures she presented as the basis of language processing by machine, now require some commentary and explanation in places if they are to be accessible today, most particularly by relating them to more recent and more widely publicised work where closely related concepts occur.

In this volume, eleven of Margaret Masterman's papers are grouped by topic, and in a general order reflecting their intellectual development. Three are accompanied by a commentary by the editor where this was thought helpful plus a fourth with a commentary by Karen Spärck Jones, which she wrote when reissuing that particular paper and which is used by permission. The themes of the papers recur, and some of the commentaries touch on the content of a number of the papers.

The papers present problems of style and notation for the reader: some readers may be deterred by the notation used here and by the complexity of some of the diagrams, but they should not be, since the message of the papers, about the nature of language and computation, is to a large degree independent of these. MMB (as she was known to all her colleagues) put far more into footnotes than would be thought normal today. Some of these I have embedded in the text, on Ryle's principle that anything worth saying in a footnote should be said in the text; others (sometimes containing quotations a page long) I have dropped, along with vast appendices, so as to avoid too much of the text appearing propped up on the stilts of footnotes. MMB was addicted to diagrams of great complexity, some of which have been reproduced here. To ease notational complexity I have in

places used 'v' and '+' instead of her Boolean meet and join, and she wrote herself that 'or' and 'and' can cover most of what she wanted. In the case of lattice set operations there should be no confusion with logical disjunction and conjunction. I have resisted the temptation to tidy up the papers too much, although, in some places, repetitive material has been deleted and marked by [...]. The papers were in some cases only internal working papers of the CLRU and not published documents, yet they have her authentic tone and style, and her voice can be heard very clearly in the prose for those who knew it. In her will she requested, much to my surprise, that I produce a book from her papers. It has taken rather longer than I expected, but I hope she would have liked this volume.

MMB would have wanted acknowledgements to be given to the extraordinary range of bodies that supported CLRU's work: the US National Science Foundation, the US Office of Naval Research, the US Air Force Office of Scientific Research, the Canadian National Research Council, the British Library, the UK Office of Scientific and Technical Information and the European Commission.

I must thank a large number of people for their reminiscences of and comments on MMB's work, among whom are Dorothy Emmet, Hugh Mellor, Juan Sager, Makoto Nagao, Kyo Kageura, Ted Bastin, Dan Bobrow, Bill Williams, Tom Sharpe, Nick Dobree, Loll Rolling, Karen Spärck Jones, Roger Needham, Martin Kay and Margaret King. I also owe a great debt to Gillian Callaghan and Lucy Moffatt for help with the text and its processing, to Octavia Wilks for the index, and to Lewis Braithwaite for kind permission to use the photograph of his mother.

YORICK WILKS
Sheffield
December 2003