Words and Intelligence II

Text, Speech and Language Technology

VOLUME 36

Series Editors

Nancy Ide, *Vassar College, New York*
Jean Véronis, *Université de Provence* and *CNRS, France*

Editorial Board

Harald Baayen, *Max Planck Institute for Psycholinguistics, The Netherlands*
Kenneth W. Church, *AT & T Bell Labs, New Jersey, USA*
Judith Klavans, *Columbia University, New York, USA*
David T. Barnard, *University of Regina, Canada*
Dan Tufis, *Romanian Academy of Sciences, Romania*
Joaquim Llisterri, *Universitat Autonoma de Barcelona, Spain*
Stig Johansson, *University of Oslo, Norway*
Joseph Mariani, *LIMSI-CNRS, France*

Words and Intelligence II
Essays in Honor of Yorick Wilks

Edited by

Khurshid Ahmad
Trinity College, Dublin, Ireland

Christopher Brewster
University of Sheffield, UK

Mark Stevenson
University of Sheffield, UK

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-5832-5 (HB)
ISBN 978-1-4020-5833-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved
© 2007 Springer
No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Contents

Biographies of the Editors vii

List of Contributors ix

Introduction xi

1. Yorick Alexander Wilks: A Meaningful Journey 1
Mark Maybury

2. Metaphor, Semantic Preferences and Context-Sensitivity 39
John A. Barnden

3. Towards a New Generation of Language Resources in the Semantic Web Vision 63
Nicoletta Calzolari

4. Information Access and Natural Language Processing: A Stimulating Dialogue 85
Robert Gaizauskas, Horacio Saggion and Emma Barker

5. Three Steps in Wilks Work: From Theory to Resources to Practice 107
Gregory Grefenstette

6. Preference Syntagmatics 119
Patrick Hanks

7. Historical Ontologies 137
Nancy Ide and David Woolner

8. An Amorphous Object Must be Cut by a Blunt Tool 153
Makoto Nagao

9. Homer, the Author of *The Iliad* and the Computational-Linguistic Turn 159
Sergei Nirenburg

10. Philosophical Engineering 195
Nigel Shadbolt

11. Machine Translation and the World Wide Web 209
Harold Somers

12. Semantic Primitives: The Tip of the Iceberg 235
Karen Spärck Jones

13. Molecules, Meaning and Post-Modernist Semantics 255
John Tait and Michael Oakes

Biographies of the Editors

Khurshid Ahmad holds the Chair of Computer Science at Trinity College, Dublin, Ireland; he was the founding Head of the Department of Computing, University of Surrey, England and held the Chair of Artificial Intelligence. He is interested in cross-modal interaction in human cognition and the simulation of such interaction in information systems. His research interests are in the areas of information extraction, neural networks, ontology and terminology systems, knowledge management and financial engineering. He has been working with major EU IT organizations and universities on problems related to terminology of specialist domains and the ontological commitment of the domain. His current projects include automatic summarisation of video streams for surveillance which involves studies of attention and distraction amongst humans; he has been working on the automatic identification of 'sentiments' in financial news reports that influence the behaviour of financial markets; and he published on multi-modal information fusion in child language and numerosity development, and on the simulation linguistic deficit amongst aphasic patients, using multi-net neural computing systems. In knowledge management he has been dealing with the diffusion of knowledge from research laboratories onto the first stage of industrialisation, patents, using a corpus based study of lexical preferences amongst scientists. He has published over 150 papers in various topics, edited two books and written one. He obtained his doctorate in theoretical nuclear physics in 1975 and has enjoyed being curious about complex systems, language and the evolution of ideas ever since. His work has been supported by the EU programmes of R&D in IT and by the UK Research Councils including the EPSRC, ESRC and AHRC. He is the Fellow of the British Computer Society.

Christopher Brewster is a Research Fellow in the Department of Computer Science at the University of Sheffield. He has worked in a number of language technology projects, having had experience especially in computational corpus linguistics and lexicography. He worked on the problem of knowledge acquisition and maintenance in the EPSRC Advanced Knowledge Technologies (www.aktors.org) contributing especially to research on ontology learning, the appropriacy of ontologies for knowledge representation, and ontology evaluation. He was lead scientist on the Abraxas project (http://nlp.shef.ac.uk/abraxas/) which focussed on ontology learning techniques. He is currently Project Manager of the Companions Project (www.companions- project.org). He has published a number of papers on the subject of ontologies, and organised several workshops.

Mark Stevenson is a lecturer and EPSRC Advanced Research Fellow at Sheffield University, where he is a member of the Natural Language Processing group. He has previously worked for Reuters Ltd. In London where he led projects on the application of language technology to business problems and also acted as the industrial contact on European projects. In 2001/2002 he was an inaugural Reuters Foundation visiting Fellow at the Center for the Study of Language and Information (CSLI), Stanford University. His research interests include word sense disambiguation, semantic similarity and information extraction and retrieval. His PhD was supervised by Yorick Wilks and explored the application of a diverse set of knowledge sources to the word sense disambiguation problem. His thesis was published by CSLI Publications. In addition, he has published around 50 papers in journals, collected volumes and international conferences.

List of Contributors

Khurshid Ahmad, Department of Computer Science, Trinity College Dublin, Dublin, Eire

Emma Barker, Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

John A. Barnden, School of Computer Science, University of Birmingham, Birmingham, United Kingdom

Christopher Brewster, Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

Nicoletta Calzolari, Istituto di Linguistica Computazionale del CNR Pisa, Italy

Robert Gaizauskas, Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

Gregory Grefenstette, CEA LIST, Fontenay aux Roses, France

Patrick Hanks, Masaryk University, Brno, Czech Republic

Nancy Ide, Department of Computer Science, Vassar College, Poughkeepsie, NY, United States of America

Mark Maybury, MITRE Corp., Bedford, MA, United States of America

Makoto Nagao, National Institute of Information and Communications Technology, Tokyo, Japan

Sergei Nirenburg, National Institute for Language and Information Technologies, University of Maryland, Baltimore County, MD, United States of America

Michael Oakes, Department of Computer Science, University of Sunderland, Sunderland, United Kingdom

Horacio Saggion, Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

Nigel Shadbolt, School of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom

Harold Somers, School of Informatics, University of Manchester, Manchester, United Kingdom

Karen Spärck Jones, Computer Laboratory, University of Cambridge, Cambridge, United Kingdom

Mark Stevenson, Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

John Tait, Department of Computer Science, University of Sunderland, Sunderland, United Kingdom

David Woolner, Marist College, Poughkeepsie, NY, United States of America

Introduction

It has been said of the brothers Wilhelm and Alexander von Humboldt that between them they were the last people to have known all that there was to know, to have had a mastery of the best that contemporary science knew and to have made significant contributions, to be that rare thing Renaissance men. In a world of ever-greater specialisation, especially in academia, the ability to cross intellectual boundaries, bring together ideas beyond the confines of one's narrow discipline and yet make significant intellectual contributions has become ever rarer. In bringing together this celebration of Professor Yorick Wilks, it has been the ambition of the editors to provide the reader with a taste, an inkling of that which cannot be conveyed on the written page but only in the person of Yorick. He is a renaissance man in an age where such concepts have been forgotten. He is a bridge between a bewildering variety of contemporary research, and simultaneously a link between some of the most advanced thought in the broadly interpreted field of Artificial Intelligence (AI) and the long tradition of philosophy, literature, and general intellectual creativity that have fundamentally informed his academic research. This comes across in part when one considers his career, more so when one reads his writings but is most apparent in person.

Modern scientists have become specialised, experts in only one specific domain. In contrast, Yorick Wilks has remained a universalist, actively contributing to a wide range of topics, from the details of machine translation or information extraction to the philosophical implications of certain AI positions or the current political situation in the world. A long history of widely cited publications in a great variety of academic organs bear testament to a highly productive and influential career which is honoured in this Festschrift and manifested in the accompanying volume of selected papers by Yorick Wilks.

Artificial Intelligence and Natural Language Processing (NLP) have been the primary areas of concern for Yorick over the years of his career and yet this has not detracted from Yorick's capacity to have competence and make contributions across a large range of topics. His academic passions have included AI, its philosophical foundations, architectures for NLP, computational syntax and semantics, lexica and lexical resources, word sense disambiguation, machine translation, knowledge representation and acquisition, belief systems and agents, human-computer dialogue and information extraction. Some, but not all of his passions, are conveyed and celebrated in the contributions contained in this volume.

During his career Yorick has led a succession of successful research groups (detailed in Mark Maybury's biographical paper). In this capacity he has a particular talent for obtaining research grants and has successfully funded large research groups in the UK and USA. At the University of Sheffield he built one of the most successful research groups in the world in the field of NLP. However, there is much more to leading such a group than writing good grant proposals. Yorick has always had a vision for his groups, a broad concept of what it was trying to achieve while concurrently allowing individual researchers freedom to pursue their own interests, to be creative. He always has confidence that if you put good people together, give them an appropriate degree of freedom, academic creativity and innovation will thrive.

What cannot be conveyed by the written word, and which we can only convey superficially in this introduction are the human qualities that have accompanied the research and learning. Yorick has a breadth both in his humanity and in his culture, a tolerance and understanding of fellow human beings, a good humour and generosity of spirit. This is apparent in the freedom he has given his students and his ability to create a fertile productive environment in order to allow research to flower. Furthermore, he embodies a sense of vision and a depth of knowledge, a deep insight into human qualities and a tolerance for human frailties, all of which are combinations both unusual and refreshing. Yorick has always conveyed passion in both his work and leisure. He has a natural ability to make people feel at ease and is a famed raconteur. Part of this comes out of Yorick's immense breadth of interests. He has always led a double life, having over the years a very successful amateur acting career (which nearly became professional according to some anecdotes). Furthermore, he is someone who in a previous generation would have been approvingly described as "well-read" and this broad culture informs both his scientific output and interests and the daily interaction he has with his colleagues. Life in AI and NLP would have been the poorer without the person of Yorick to bring his *joie de vivre*.

It has always been a privilege to be a research student supervised by Professor Wilks. Yorick is both visionary and practical, encouraging the student to read a text, whether from last year's conference or a hundred years ago, and place it in the context of their current concerns. It is the content of the ideas that interest him in a student's work without any concern for formality or procedure. And this is another area where Yorick's capacity for seeing the potential of people is most apparent, a potential which they will usually not be aware of themselves. Above all Yorick has been able to create an environment in which one is encouraged to publish, attend conferences and carve out an independent research career. As much as he can Yorick has always sought to support students through their studies, financially and intellectually. It was in this spirit of developing students that Yorick was instrumental in founding the CLUK (Computational Linguistics UK) series of conferences focussed on the needs of graduate students in the UK.

One of Yorick's strength is his ability to collaborate with academics and end-users with different interests, in synthesising complex ideas across different disciplines, and finally in articulating such ideas. There are many instances of such

collaborations across the world, many reflected in the contributions in this Festschrift. Collaborating with Yorick has always been an extremely stimulating experience right from the start of the conception of the project, through the writing of proposal texts (in which Yorick excels), to the realisation of the project and its outcomes in collaboratively written papers or software. His capacity to bring together collaborators from different countries, different cultures, let alone entirely different academic communities is one much celebrated and recognised.

The collection begins with a biographical essay on Yorick written by Mark Maybury. Maybury has collected details and anecdotes from a wide circle of Yorick's friends and acquaintances to present an amusing and insightful account of a life full of meaning in all senses. In his paper entitled "Metaphor, Semantic Preferences and Context-Sensitivity," John Barnden discussed Yorick's work on metaphor which he interprets as being "utterance-based" while arguing for a context-based approach. Nicoletta Calzolari, in her paper "Towards a new generation of Language Resources in the Semantic Web vision", notes Yorick's early and prophetic understanding of the importance of natural language corpora, and while reviewing a number of language resources related projects in which Yorick has been involved, makes the case for a continuing need for infrastructure focussed HLT research.

Robert Gaizauskas, a long time colleague of Yorick's at the University of Sheffield, has contributed a paper co-authored with Emma Barker and Horacio Saggion on "Information Access and Natural Language Processing: A Stimulating Dialogue." This considers the role of NLP in relation to IR and information access in general with specific reference to a project "Cub Reporter" undertaken at the Sheffield NLP group. Gregory Grefenstette's paper on "Three steps in Wilks work: From theory to resources to practice" is a celebration of what the author sees as three important components in Yorick's work viz. his "flight's of brilliance," "his reasoned response to difficulties" and his unrelenting engineering effort and serious science. Grefenstette considers in detail three papers of Yorick's from what could be called the early, middle and late work, placing the work in the context of contemporary research.

In a paper entitled "Preference Syntagmatics" Patrick Hanks discusses an ongoing project of his to create a "Pattern Dictionary" which is fundamentally influenced by Yorick's early work on Preference Semantics. A central claim is that the problem of word sense disambiguation needs to be reformulated before it can successfully be resolved. On a somewhat different tack "Historical Ontologies", a paper by Nancy Ide in collaboration with David Woolner, discusses the challenge of creating knowledge representations that can handle diachronic events.

A major interest of Yorick has been machine translation and Makoto Nagao in his paper "An Amorphous Object Must Be Cut By A Blunt Tool" gives an account of the creation of example based machine translation, an approach he claims is typically Japanese. Addressing another ongoing concern of Yorick's is Sergei Nirenburg's paper "Homer, the Author of The Iliad and the Computational-Linguistic Turn." He sets out in detail the disagreement, first, between Yorick's views on knowledge representation and the tradition of Fodor, and then Yorick's views on knowledge representation resources such as ontologies and the views of the Guarino.

Reflecting the wider impact Yorick has had as a philosopher concerned with AI and computer science, Nigel Shadbolt's paper "Philosophical Engineering" discusses the fundamental philosophical issues which arise in undertaking modern computer engineering. He notes that while formal models have immense power in a continuously changing world our capacity to construct models is under constant challenge. Returning to machine translation, Harold Somers gives a brief account of Yorick's long term involvement and impact on the field, and goes on to consider current machine translation freely available on the web, and its success and impact.

Yorick's near contemporary at Cambridge, Karen Spärck Jones has contributed a paper on "Semantic primitives: The tip of the iceberg" which discusses how semantic primitives, a long time concern of Yorick's, are considered today. The collection concludes with a paper by John Tait in collaboration with Michael Oakes "Molecules, Meaning and Post-Modernist Semantics" which again return to preference semantics but from the perspective of the need for machine learning of lexical resources.

The editors have put together this Festschrift, and the accompanying volume of Selected Papers, in order to celebrate him as an individual and bring into focus his work and its impact across a range of research topics. We would like to thank all the contributors to this volume for taking time out of their schedules to write these papers and thus make the Festschrift possible. We have known Yorick in various capacities, as students, colleagues, collaborators and friends, and we sincerely hope that these volumes will bring pleasure to him, his colleagues and friends.

Khurshid Ahmad
Christopher Brewster
Mark Stevenson

Acknowledgements

The editors would like to thank Springer Verlag for producing these two volumes and also Nancy Ide whose dinner invitation at LREC in Lisbon and conversations with Jolanda Voogd led to these volumes.