Information Sharing:
Reference and Presupposition in Language Generation and Interpretation

Kees van Deemter and Rodger Kibble

ISBN: 1-57586-404-5

Contents

Contributors

Antonella De Angeli LORIA Laboratory, Vandœuvre-lès-Nancy, France. Antonella.De_Angeli@ncr.com

John Bateman Angewandte Linguistik, Universität Bremen, Germany. bateman@uni-bremen.de

Norman Creaney University of Ulster, Coleraine, N. Ireland, UK. n.creaney@ulst.ac.uk

Cassandre Creswell Department of Linguistics, University of Pennsylvania, Philadelphia, USA. creswell@ling.upenn.edu

Kees van Deemter Information Technology Research Institute, University of Brighton, U.K. Kees.van.Deemter@itri.brighton.ac.uk

John Gurney Adelphi Laboratory Center, Adelphi, Maryland, USA. gurney@arl.army.mil

Herman Hendriks Utrecht Institute of Linguistics (OTS), Utrecht University, The Netherlands; and ILLC/Department of Philosophy, University of Amsterdam, The Netherlands. Herman.Hendriks@let.uu.nl

Pamela W. Jordan Learning Research and Development Center University of Pittsburgh, Pennsylvania, USA. pjordan@pitt.edu

Rodger Kibble Department of Mathematical and Computing Sciences, Goldsmiths College, University of London, U.K. r.kibble@gold.ac.uk

Elizabeth Klipple Adelphi Laboratory Center, Adelphi, Maryland, USA. klipple@arl.army.mil

Emiel Krahmer Computational Linguistics, Tilburg University, The Netherlands. e.j.krahmer@kub.nl

Geert-Jan Kruijff Computerlinguistik, Universität des Saarlandes, Saarbrücken, Germany. gj@coli.uni-sb.de

Ivana Kruijff-Korbayová Computerlinguistik, Universität des Saarlandes, Saarbrücken, Germany. korbay@coli.uni-sb.de

Frédéric Landragin LORIA Laboratory, Vandœuvre-lès-Nancy, France. Frederic.Landragin@loria.fr

Patrice Lopez LORIA Laboratory, Vandœuvre-lès-Nancy, France. Patrice.Lopez@loria.fr

Ivandré Paraboni Information Technology Research Institute, University of Brighton, U.K. Ivandre.Paraboni@itri.brighton.ac.uk

Craige Roberts Department of Linguistics, Ohio State University, Columbus, Ohio, USA. croberts@ling.ohio-state.edu

Laurent Romary LORIA Laboratory, Vandœuvre-lès-Nancy, France. Laurent.Romary@loria.fr

Jennifer Spenader Computational Linguistics, Department of Linguistics, Stockholm University, Sweden. jennifer@ling.su.se

†**Rosemary Stevenson** Department of Psychology, University of Durham, U.K.

Mariët Theune Parlevink Language Engineering Group, Department of Computer Science, University of Twente, Enschede, The Netherlands. theune@cs.utwente.nl

Frédéric Wolff LORIA Laboratory, Vandœuvre-lès-Nancy, France. Frederic.Wolff@fr2.vdogrp.de

Henk Zeevat Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Netherlands. henk.zeevat@hum.uva.nl

Preface

Academic communities are sometimes separated by artificial boundaries. The linguistic community is no exception: some linguists insist on a theoretical perspective, others prefer a data-oriented or computational approach. It has often been observed that boundaries between disciplines or scientific approaches can become counterproductive if they become rigid; the present volume is an attempt to show what happens when they are taken with a pinch of salt. The papers contained in this volume might be partitioned into the three categories mentioned above, but such a partitioning would tend to obscure existing commonalities. In our view, the most important common element is what we will call *Information Sharing*.

When we express ourselves, we share information. Information sharing involves distinguishing between parts of an utterance that express *given* information (e.g., because the information has been shared before) and parts that express *new* information. Information sharing is crucial to all the papers collected in this book, regardless of their methodology. And even if we look at specific aspects of Information Sharing, such as salience, presuppositionhood, or deixis, the same picture emerges: A reader interested in how *salience* influences the form and position of a referring expression, for example, could consult papers representing each of the three types: she could read the chapters by Hendriks (theoretical), Stevenson, Jordan, and Landragin (data-oriented), or those by Kruijff et al., or Krahmer & Theune (mainly computational). A similar picture would emerge for someone interested in *deixis*: such a person would be equally likely to turn to Roberts, or to Gurney & Klipple (mainly theoretical) as to Paraboni & van Deemter (mainly computational) or Landragin et al. (data-oriented). *Presuppositions* take up a slight different position in this volume, since they are mainly studied from a theoretical perspective (Zeevat) or a data-oriented one (Spenader), but the reader

will have little difficulty finding computational approaches elsewhere (e.g., Johan Bos (2001), "DORIS 2001: Underspecification, Resolution and Inference for Discourse Representation Structures", In Proceedings of ICoS-3, Inference in Compuational Semantics.)

A few additional observations may be worth reporting. Firstly, the book contains a large number of papers that address linguistic issues from the perspective of language *generation*. A focus on language generation is manifest not only in each of the computationally oriented chapters, but also in those by Stevenson, Jordan, Landragin et al., and Creswell. Language *interpretation* has long been the dominant perspective in semantics; yet, a majority of the chapters of this volume represent the opposite perspective, asking how and under what circumstances a particular kind of expression can be produced, or how a particular kind of information is best expressed. In generation, the distinction between given and new information is as crucial as it is in language interpretation. This is evident in the generation of referring expressions, for example: when a speaker wants to say something about an object, she has to use properties whose extensions are known by the hearer (i.e., *given* properties), in order to distinguish the target from a set of contextually available (i.e., given) 'distractors'. The distinction between given and new lies at the heart of all algorithms in this area, but the papers in this volume (most notably those by Krahmer and Theune and by Creaney) elaborate on it in nontrivial ways, by making givenness a graded notion (i.e., incorporating degrees of salience) and by applying the distinction between givenness and novelty to quantified noun phrases.

Secondly – and relating to the theme of givenness and novelty in a different way – the book counts a large number of new departures. Some of the linguistic phenomena addressed here have received little attention in the literature. This is true, for example, for the work on deixis to properties (Gurney and Klipple: 'I'd like to fly *that* high'), and for the work on emphatic reflexives (Creswell: 'The president *herself* led the discussion'). It is also true for Paraboni and van Deemter's study of document deixis (as in 'The algorithms in *the concluding section of this paper*'), and for Creaney's investigation of the computational generation of quantified noun phrases. In other cases it is mainly the methodology that is new. This holds, for example, for Spenader's chapter, which confronts Van der Sandt's theory of 'presuppositions-as-anaphors' with corpus data, and it is equally true for Jordan's corpus study of referring expressions, as well as Zeevat's application of Optimality Theory to the generation of presupposition triggers. In these latter cases, the subject of study is kept constant, but it is precisely the above-mentioned methodological perspective that the authors bring to bear on this particular subject (i.e.,

theoretical, computational, or data-oriented) that is novel. Although a majority of the chapters focus on the study of nominal expressions, the present collection contains a number of papers (most notably those by Spenader, by Zeevat, by Klipple & Gurney, by Kruijff et al., and by Landragin et al.) that set their eye beyond the noun phrase. The results are interesting, we believe, and we hope that they will prove seminal.

The idea for this book came to us after the 11th European Summer School 'Logic Language and Information' (ESSLLI) in Utrecht in 1999, when three research workshops turned out to be remarkably similar in their scope and purpose. One of the three, which was concerned with the *Generation of Nominal Expressions*, was organized by us. We are grateful to Elisabeth André, Massimo Poesio, and Hannes Rieser (who organized the workshop on *Deixis*), and to Bart Geurts, Manfred Krifka, and Rob van der Sandt (who organized the workshop on *Focus and Presuppositions*). Without their efforts, this book would have lacked its present breadth. Particular thanks are due to Bart Geurts and Paul Piwek for their editorial advice. In addition, we believe that many of the virtues that we would like to see in this volume can be traced back to the multidisciplinary spirit of the yearly ESSLLI, which has grown into so much more than just a summer school. We thank the people who reviewed the papers in this volume during any stage of their bibliographic lives: when submitted to one of the workshops, when submitted for inclusion in the book proposal, and during review of the book itself. In particular (and not counting reviewers who are also contributors to this book), we thank *David Beaver, Robert-Jan Beun, Daniel Bűring, Lynne Cahill, John Carroll, Hua Cheng, Judy Delin, Christy Doran, Miriam Eckert, Bart Geurts, Jonathan Ginzburg, Tony Hartley, Petra Hendriks, Renate Henschel, Ruth Kempson, John Lee, Alice ter Meulen, David Milward, Jon Oberlander, Daniel Paiva, Paul Piwek, Massimo Poesio, Richard Power, Ehud Reiter, Rob van der Sandt, Mark Steedman, Matthew Stone, Michael Strube, Marc Swerts,* and *Jacques Terken* for their reviews.

We hope that, by sharing the information in this volume with CSLI Publications' readership, the different approaches to Information Sharing are brought a small step closer to integration.

Kees van Deemter and Rodger Kibble.