Briefly Noted

The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues

Paola Merlo and Suzanne Stevenson (editors)
(University of Geneva and University of Toronto)

Amsterdam: John Benjamins Publishing Company (Natural language processing series, edited by Ruslan Mitkov, volume 4), 2002, viii+362 pp; hardbound, ISBN 90-272-4987-3 and 1-58811-156-3, $100.00, €110.00

This volume contains 15 papers from a special session held at the 11th CUNY Sentence Processing Conference (1998). There are three thematically arranged sections: proposals from theoretical linguistics, the lexicon–syntax relationship as regards processing, and the content of lexical entries. The first chapter summarizes the papers and discusses the broader issues raised in the book. (The full list of the papers is available from the publisher's Web page.)

There are many interesting and original contributions from computational and behavioral perspectives. These shed new light on open questions such as modularity in sentence processing, the distinction between lexical and structural processing, and argument structure relations with respect to verbal lexical entries.

This book is an excellent resource for researchers interested in human language processing, and the computational linguist interested in the connection between probabilistic parsing and human sentence processing will find a number of the papers very satisfying.—*Shravan Vasishth, Universität des Saarlandes*

Virtual Humans: A Build-It-Yourself Kit, Complete with Software and Step-by-Step Instructions

Peter Plantec

New York: Amacom (American Management Association), 2004, xviii+279 pp and CD-ROM; paperbound, ISBN 0-8144-7221-4, $34.95

Most researchers of **embodied conversational agents** (ECAs), or as Peter Plantec prefers to call them, virtual humans (V-humans), know only too well that there's nothing that annoys a computational linguist quite as much as an interactive animated conversational character. This is especially true if regular people really enjoy playing with it—as they often do. For the traditionalist, the use of a graphical character is a cheat, a conjuror's misdirection aimed at fooling the naive user.

This book will do little to ease the aggravation. Indeed its folksy psychology and sweeping characterization of linguistics and cognitive psychology will only heighten the sensation. This said, academic and nonacademic books on ECAs are few and far between, and here Plantec demonstrates that there is rather more to his notion of a V-human than a few hastily composed regular expressions. This was never going to be a book on computational linguistics but is rather a guide to faking personality, and for this Plantec is unapologetic. While his science is suspect, it's easy to argue that his focus on personality is well founded, as recent funding initiatives in cognitive and computational studies of emotion demonstrate.

Describing the contents of a book like this is a challenge too far. Plantec skips from the V-human market to a characterization of consciousness and synthetic thinking, all in the first three chapters. Much of the remainder of the book is peppered with informal accounts of the syntax and capabilities of particular pattern-matching formats and animation schemes. While there's little coherence, there is a genuine wealth of information as to the art of building engaging patterns of conversation and graphical behavior.

Plantec's accounts of different V-human technologies (most of which are in fact rather accessible) are limited to chatterbots but still are of practical value and could serve as a useful introduction to anyone very new to ECAs and interested in quickly building some kind of system. Personally, I would have bought the book for the accompanying CD, which includes samples of graphical authoring tools, chatterbots, and a few lovely pieces of video and audio footage of past systems.

In the book's introduction, Ray Kurzweil predicts ECAs with humanlike capabilities by the 2030s. Of course, all this tells us is that he's probably not expecting to be around then. Chatterbots coupled to Microsoft Agent are not the future of ECAs, and this imper-

fect account of their moderate success (if only in engaging the casual user) is only possibly worth a second look. However, it very probably is a great starting point for some enjoyable undergraduate projects.—*Patrick Olivier, Lexicle*

Semantic Relations and the Lexicon: Antonymy, Synonymy, and Other Paradigms

M. Lynne Murphy
(University of Sussex)

Cambridge, England: Cambridge University Press, 2003, ix+292 pp; hardbound, ISBN 0-521-78067-5, $65.00

An understanding of semantic relations, such as hyponymy and synonymy, is important in many applications in natural language processing, and typically such relations are predetermined and listed in the system's lexicon or other "lexical resource." WordNet is the resource most commonly used. Murphy's view is that such relations are not lexical but "metalexical"—stored not in the mental lexicon but at the next level up—because they are knowledge *about* words, not *of* words; they are not arbitrary, nonderivable information for linguistic competence. Moreover, they are context-dependent.

Murphy thus argues for what she calls a pragmatic view of semantic relations, which she bases on her principle of **relation by contrast**: "The contrast relation holds among the members of a set iff they have all the same contextually relevant properties but one" (page 44). In defense of this position, Murphy gives an extensive critique of other theories of semantic relations, draws on both linguistic and psycholinguistic evidence to support her case, and presents in-depth treatments of synonymy and of contrast relations such as antonymy.

Murphy's critique of WordNet as static and context-free is familiar in computational linguistics. But the alternative is not obvious. One can readily agree with Murphy that semantic relations are context-dependent, but the problem of actually determining the relations computationally in context is beyond the current state of the art; Murphy's analysis does not include any suggestions for algorithmic methods for deriving relations. Moreover, the interaction between semantic relations and context may be two-way: It can be the relations that construct the context.

While Murphy's treatment of semantic relations is not at all computational, her perspective and the problems that arise from her work will be interesting to any researcher of computational problems in lexical semantics.—*Graeme Hirst, University of Toronto*

Ellipsis and Reference Tracking in Japanese

Shigeko Nariyama
(University of Melbourne and Nara Institute of Science and Technology)

Amsterdam: John Benjamins (Studies in language companion series, edited by Werner Abraham and Michael Noonan, volume 66), 2003, xv+397 pp; hardbound, ISBN 90-272-3076-5 and 1-58811-420-1, $150.00, €125.00

"There is no grammatical requirement for nominal arguments, such as subject and object, to be overt in Japanese, and they are frequently unexpressed. . . . The main aim of this book is to elucidate the linguistic mechanisms whereby the referents of unexpressed arguments are identified. . . . These unexpressed arguments, referred to as 'ellipsis' in this book, are formidably prevalent in Japanese. . . . Despite this high frequency of ellipsis, Japanese is not equipped with the grammatical devices such as cross-referencing systems and verbal inflections commonly found in pro-drop languages for referent identification, and yet the mechanisms of ellipsis remain little explicated in the literature. Hence, it is the general view in the current literature that the referential identification of ellipsis is sought in 'context', 'mutual knowledge', 'sociolinguistic variables', or even 'intuition'. . . .

"The mechanisms for referential identification of ellipted arguments are extremely intricate but by no means as elusive as the commonly held view would have it. No matter how vague or illogical Japanese may look to non-native speakers of Japanese, since speakers of Japanese have little trouble communicating among themselves, there must be concrete linguistic mechanisms that they use to identify the referents of ellipted arguments. . . . In terms of machine translation systems from and into Japanese, however, because machines do not have the faculty that humans do and cannot rely on in-

tuition as such, they require explicit procedures for deducing the referents of ellipted arguments. The major goal of this book is to provide concrete and precise procedures that are not only comprehensible to humans but are also suitable for machine translation systems. . . . These mechanisms stem from three tiers of linguistic information: first, the use of argument-inferring morphemes on verbal predicates; secondly, tightly constrained argument structures which anchor the subject and which induce patterns of ellipsis; and thirdly, cohesively sequenced discourse structures with the topicalized subject as pivot and the distinctions marked by the use of *wa* and *ga*. The interplay among these three tiers of morphology/semantics, syntax, and discourse devices is the key to determining the referential identity of ellipted arguments. I devise an algorithm summing up these mechanisms, and demonstrate, using naturally occurring texts, how it can both detect the existence of ellipsis in sentences and track referential identity." —*Abridged from the author's introduction*