Briefly Noted

Word Manager: A System for Morphological Dictionaries

Marc Domenig and Pius ten Hacken
(Universität Basel)

Hildesheim, Germany: Georg Olms Verlag (Informatik und Sprache 1), 1992, ix + 211 pp.
Paperbound, ISBN 3-487-09677-3, no price listed

This book describes Word Manager (WM), a lexical database system intended to serve a wide range of NLP applications. The most striking aspect of WM itself, and of this presentation, is its windowed user interface. The book consists of five chapters ('Introduction,' 'The Word Manager Approach,' 'Inflection,' 'Wordformation' [sic], and 'Linguistic Theories') and a lengthy appendix giving syntax definitions for the user language.

A WM linguistic description consists of a tree-shaped hierarchy whose nodes contain declarations, feature specifications, and rules and entries of various kinds. Each node is identified by a feature set and associated with a window, the label of the node being inherited by those below it in the tree.

Affixation in its simplest form is indicated by juxtaposing feature specifications that identify the segments concerned. More complex cases requiring spelling adjustments at segment boundaries are treated by means of 'match and map' rules. These are regular expression pattern matchers with bindings and substitutions, and may be paired with feature sets that govern their application. WM spelling rules differ from the more familiar two-level variety in being ordered. So, for example, the plural form *amici* of the Italian *amico* 'friend' can be produced by adding *h* to the stem (as would be done for the normal case *baco, bachi* 'silkworm') and then removing it again. The authors adopt without comment extrinsic rule ordering and radical non-monotonicity when elsewhere the tendency, shared by many theoretical linguists, has for some time been to abandon them.

One reason for choosing a hierarchical organization is that redundancy can be minimized through the use of inheritance. However, several aspects of WM give rise to unnecessarily redundant specifications. Patterns in spelling rules must apparently subsume the entire string against which they match; this leads to the presence of repeated subexpressions whose only function is to skip irrelevant characters. What has been missed here is the fact that the phenomena these rules are intended to handle are essentially boundary effects. The pattern-matching component of the rules shows no sign of having been designed for, or even adapted to, the purpose for which it is employed. Similarly, it seems necessary to specify for each word segment mentioned in a rule not only its lexical form but also all of its surface forms, in addition to supplying spelling rules that implicitly express the same correspondence. The intention is to permit cross-checking during compilation, but the tracing facilities offered by WM should make this unnecessary. There is a general impression of piecemeal design, almost as if the shell of WM had been developed without regard to linguistic considerations and then fleshed out with rules and entries at the last moment when it was too late to change anything.

But the real weak point of this book lies less in WM itself than in the presentation. Typically, this proceeds by describing some morphological phenomenon, and comparing two or three possible analyses. The emphasis is entirely on examples and the syntax of the system; nowhere do we find a clear statement of how the syntax is to be interpreted, an account of the formal properties of the various mechanisms employed, or proper motivation for the choice of these mechanisms rather than others. There are some interesting ideas hidden below the surface (morphological rules can be specialized to handle exceptions, feature values are used to encode paths through the hierarchy), but it is hard to evaluate them in this form.

The first chapter sets out to provide justification for WM and draws comparisons with other approaches to morphology and lexical organization. The authors' awareness of such work seems quite rudimentary: the three contrasted approaches are finite-state lexicons with two-level rules (Koskenniemi 1983), DATR (Evans and Gazdar 1990), and the Celex databases. No mention is made of, for example, the lexical knowledge base created for the Acquilex project (Copestake [1992] is a recent report, but the project has been well publicized for much longer), or the large amount of work that has been done by Bear (1988), Trost (1991), and oth-

ers to augment the original two-level method with more powerful devices. The omissions would be less damaging had the authors not couched the discussion in terms of an evaluation of competing proposals from which WM emerges as the victor. The chapter as a whole is marked by a distastefully patronizing tone, and contains many unsupported and contentious statements.

Finally, it seems to me that the authors have greatly overestimated the utility of their approach. The central idea is that WM will provide a purely morphological resource for a wide range of applications, including MT, spelling correctors, text-to-speech, and so on, in a client-server arrangement. Most of these require other kinds of information (phonological, semantic, translation, etc.), which WM provides no natural way of expressing. As a result, in order to support any but the most trivial application, a parallel dictionary must be created and accessed alongside that available via WM. Contrary to the authors' claim, this is not simply a matter of building an interface; it is likely often to require at least as much work as creating the original dictionary. Given the network context they assume, maintaining consistency between the two sets of data will impose an extra burden.

The book is a strange mixture, and not one that I can recommend. Some parts of it (like the desultory and shallow "evaluation of other approaches") are reminiscent of a dissertation, while others (annotations to the appendix and a description of some bizarre conventions concerning the interpretation of white space) read like a satirical parody of a user manual; interspersed with the surrounding system description is a naïve and unconvincing sales pitch.—*Graham Russell, ISSCO, University of Geneva*

References

Bear, John (1988). "Morphology with two-level rules and negative rule features." In *Proceedings, 12th International Conference on Computational Linguistics*, Budapest, August 1988, 28–31.

Copestake, Ann (1992). "The Acquilex LKB: Representation issues in semi-automatic acquisition of large lexicons." In *Proceedings, 3rd Conference on Applied Natural Language Processing*, Trento, April 1992, 88–95.

Evans, Roger, and Gazdar, Gerald (1990).
"The DATR Papers." Cognitive Science Research Paper CSRP 139, School of Cognitive and Computing Sciences, University of Sussex.

Koskenniemi, Kimmo (1983). *Two-level morphology: A general computational model for word-form recognition and production.* Publication 11, Department of General Linguistics, University of Helsinki.

Trost, Harald (1991). "X2MORF: A morphological component based on augmented two-level morphology." In *Proceedings, 12th International Joint Conference on Artificial Intelligence*, Sydney, August 1991, 1024–1030.

Arenas of Language Use

Herbert H. Clark
(Stanford University)

For many years, Herbert Clark has studied the pragmatics of how people actually use language and how reality differs from neat linguistic theory. His research brings together the methods and theories of experimental psycholinguistics, sociolinguistics, and intentional analyses of language; and a knowledge of Clark's research and his results is important for anyone in computational linguistics who wants to make computers use language in the same way that people do. Many readers of this journal will already be familiar with Clark's work; those who aren't will find *Arenas of Language Use* to be a convenient introduction.

The book contains 12 papers by Clark and his students and colleagues, all but one of them previously published (between 1981 and 1991), that weave together, in different ways, Clark's central themes: the crucial role in language understanding of common ground and mutual knowledge among conversation participants; the collaborative nature of conversation; the difference between merely hearing a conversation and participating in it; and the methods by which conversants construct and understand referring expressions.

Almost all the papers remain up-to-date and relevant. (The only exceptions are the

two 1983 chapters on nonce words, largely superseded by recent research on metonymy and metaphor.) All have been re-typeset for this volume (with just a few obvious typos and missing references). There is much in this book that computational linguistics can learn from.—*Graeme Hirst, University of Toronto*

Linguistic Issues in Machine Translation

Frank Van Eynde (editor)
(Belgian Fund for Scientific Research and University of Leuven)

London: Pinter Publishers (Communication in Artificial Intelligence Series, edited by Robin P. Fawcett and Erich Steiner), 1993, viii + 239 pp.
Distributed in North America by
St. Martin's Press
Hardbound, ISBN 1-85567-024-0, $79.00

What a welcome addition to the collection of books on machine translation research this book is! Although this volume of six articles spans an eclectic mix of linguistic issues and theoretical approaches, this range reflects the current refreshing burst of energy in the field.

Aficionados of lexical-functional grammar can jump to Sadler's chapter for one view of the broader ongoing debate to define both the strengths and limits of this linguistic theory as extended to MT research. (For another recent contribution to the debate, see Kaplan and Wedekind 1993.) The chapter also contains several "translationally relevant" arguments in favor of shifting the LFG-based computational approach from one that is "structural" (or recursive) to one that is constraint-based.

By contrast, those interested in dependency theory can turn to Badia's chapter. There the dependency-theory concepts of governor and dependency relations provide the theoretical framework for one level of representation in MT systems, namely the "interface structures" that encode the source or input language sentence at the final stage of analysis, just before translation per se begins in a transfer-based system. The semantic predicative relations presented here, as well as their limitations, are well known to MT researchers (see Hutchins and Somers 1992). What will be intriguing to those involved in interlingua research is the particular way that relations traditionally captured in inter-lingual structures have been integrated into the transfer-based approach.

Another view of MT research comes from the chapters that address the general requirements that any MT system must meet if it is to handle the translation of specific types of linguistic data: unbounded dependencies, prepositions, and the subparts of words. These chapters, by Allegranza, Durand, and Bennett respectively, lean more extensively on the reader's background knowledge of general linguistics than do the two theory-specific chapters mentioned above. Of all the research challenges raised by these three authors, Bennett has identified the one that is still the most neglected: developing principles for the compositional translation of subparts of words. This shows up, as it were, close to home: though Durand does not focus at this level in his chapter, he too will eventually have to work on translating the subparts of words—prepositions in one language can translate to affixes in other languages.

Finally a word about the editor's chapter in the book: Van Eynde holds the uncontroversial position that transfer-based systems are too difficult and costly to scale up, and furthermore, that their design is not optimized for reusability. His design requirements for addressing these issues are that transfer-based MT systems must be cleanly divided into monolingual and bilingual modules, and that the functions of the bilingual modules should be reduced as far as possible. The most fascinating part of this chapter is Van Eynde's proposal for this reduction: the closed-class elements are to be represented interlingually and removed from the bilingual modules. Perhaps Van Eynde's proposal will inspire even the staunchest of interlingua proponents to take a new look at transfer-based MT research.—*Clare R. Voss, University of Maryland*

References

Hutchins, John W., and Somers, Harold L. (1992). *An Introduction to Machine Translation*. London: Academic Press.

Kaplan, Ronald M., and Wedekind, Jürgen (1993). "Restriction and correspondence-based translation." In *Proceedings, Sixth Conference of the European Chapter of the Association for Computational Linguistics*, Utrecht, The Netherlands, 193–202.

Formal Semantics: An Introduction

Ronnie Cann
(University of Edinburgh)

Cambridge, England: Cambridge University Press (Cambridge Textbooks in Linguistics, edited by J. Bresnan et al.), 1993, xvii + 344 pp.
Hardbound, ISBN 0-521-37463-4, $69.95;
Paperbound, ISBN 0-521-37610-6, $22.95

This book is an introduction to formal semantics, especially Montague semantics, within a linguistic framework. It presupposes no previous background in logic, but takes the student from simple predicate–argument structures and their interpretation through to Montague's intensional logic, covering set theory, propositional logic, type theory, lambda abstraction, traditional and generalized quantifiers, inference, tense and aspect, possible-worlds semantics, and intensionality. The emphasis throughout is on the use of logical tools for linguistic semantics, rather than on purely logical topics, and the introductory chapter situates formal semantics within the general framework of linguistic semantics. The book assumes some basic knowledge of linguistics, but aims to be as nontechnical as possible within a technical subject.—*Adapted from the publisher's announcement*

The Elements of Mathematical Semantics

Maurice V. Aldridge
(University of the Witwatersrand)

Berlin: Mouton de Gruyter (Trends in Linguistics: Studies and Monographs, edited by Werner Winter, Volume 66), 1992, xi + 261 pp.
Hardbound, ISBN 3-11-012957-4, DM 148.00

Similar in many ways to Cann's book (see previous note), *The Elements of Mathematical Semantics* works within the Montague framework and categorial grammar. Topics covered include vagueness and ambiguity, logical form in Chomskyan binding theory, presupposition and other aspects of pragmatics. The book presupposes no previous knowledge of set theory or formal logic.

Talking Data: Transcription and Coding in Discourse Research

Jane A. Edwards and Martin D. Lampert (editors)
(University of California, Berkeley)

Hillsdale, NJ: Lawrence Erlbaum Associates, 1993, viii + 325 pp.
Hardbound, ISBN 0-8058-0348-3, $59.95;
Paperbound, ISBN 0-8058-0349-1, $27.50

This book is intended as a handbook for established researchers who are interested in preparing transcribed and/or coded data or locating and using transcribed and coded data of others. It presents the reader with a set of diverse, carefully developed, and clearly specified systems of transcription and coding, arising from contrasting theoretical perspectives, and presented as alternative choices, situated within the theoretical domain most natural to each. The perspectives represented in the book include first and second language acquisition, interethnic and cross-cultural interaction, information structure, and the study of discourse influences on linguistic expression.—*Adapted from the editors' preface*

Theorie und Praxis des Lexicons [Theory and Practice of the Lexicon]

Frank Beckmann and Gerhard Heyer (editors)
(Ruhr-Universität Bochum and Truimph/Adler, Nürnberg)

Berlin: Walter de Gruyter (Grundlagen der Kommunikation und Kognition, edited by Roland Posner and Georg Meggle), 1993, viii + 348 pp.
Hardbound, ISBN 3-11-013502-7, DM 198.00

This book is based on a workshop held in Bochum in November 1992. The contents are as follows:

Frank Beckmann, "Theorie und Praxis des Lexikons. Einleitung [Theory and practice of the lexicon. Introduction.]"
Renate Bartsch, "What is in the (mental) lexicon?"
Petr Sgall, "The role of lexical data in a dependency-based description"
Dieter Wunderlich, "Funktionale Kategorien im Lexikon [Functional categories in the lexicon]"
Horst Singer, "Sprachwissen und Weltwissen im Lexikon [Linguistic

knowledge and world knowledge in the lexicon]"

Peter Bosch, "Bemerkungen zu den empirischen Grundlagen lexikalischer Semantik [Remarks on the empirical foundations of lexical semantics]"

Christoph Schwarze, "Primäre und sekundäre Lokalverben im Französischen [Primary and secondary position verbs in French]"

Udo L. Figge, "Konzeptsystem als Grundlage fur die Lexikographie [Conceptual system as the basis for lexicography]"

Gert Rickheit and Hans Strohner, "Zu einer kognitiven Theorie konzeptueller Inferenzen [On a cognitive theory of conceptual inference]"

Willem J. M. Levelt, "Lexical selection, or How to bridge the major rift in language processing"

Reinhard Köhler and Gabriel Altmann "Begriffsdynamik und Lexikonstruktur ['Begriffsdynamik' and structure of the lexicon]"

Wolfgang Klein, "Some notorious pitfalls in the analysis of spatial expressions"

Gerhard Heyer, "On the role of the dictionary and dictionary-based approaches in language products technology"

Maurice Gross, "Lexicon based algorithms for the automatic analysis of natural language"

Siegfried Kanngießer, "Zur Wiederverwertung lexikalischer Information [On recycling lexical information]"

Rebecca Bruce, Louise Guthrie, and Yorick Wilks, "Automatic lexical extraction—Theories and applications"

Nicoletta Calzolari, Johan Hagman, Elisabetta Marinai, Simonetta Montemagni, Antonietta Spanu, and Antonio Zampolli, "Encoding lexicographic definitions as typed feature structures"

Wolf Paprotté, "Typed Feature Structures und die Repräsentation flexionsmorphologischer Information im Lexikon [Typed feature structures and the representation of inflectional morphology in the lexicon]"

Corpus-Based Computational Linguistics

Clive Souter and Eric Atwell (editors)
(University of Leeds)

This book contains papers presented at the 12th Conference of the International Computer Archive of Modern English, Leeds, May 1991. The contents are as follows:

Laurie Bauer, "Progress with a corpus of New Zealand English and some early results"

Sidney Greenbaum, "The tagset for the International Corpus of English"

Charles Meyer and Richard Tenney, "Tagger: An interactive tagging program"

Lou Burnard, "The Text Encoding Initiative: A further report"

Gerry Knowles, "From text to waveform: Converting the Lancaster/IBM Spoken English Corpus into a speech database"

Richard Piepenbrock, "A longer term view on the interaction between lexicons and text corpora in language investigation"

Anne Wichmann, "Gradients and categories in intonation: A study of the perception and production of falling tones"

Josef Schmied, "Qualitative and quantitative research approaches to English relative constructions"

Christine Johansson, "*Whose* and *of which* with nonpersonal antecedents in written and spoken English"

Jacques Noël, "Using the COBUILD examples to improve the grammar notes"

Kay Wikberg, "Verbs as indicators of text type and/or style: Some observations on the LOB Corpus"

Pieter de Haan, "Sentence length in running text"

Paul Gorman and Nigel Hardy, "CLAWS, Ada, and software components"

Clive Souter, "Harmonising a lexical database with a corpus-based grammar"

Eric Atwell, "Corpus-based statistical modelling of English grammar"

Andrew Wilson and Paul Rayson,
  "Automatic content analysis of spoken
  discourse"
Louise Guthrie, "A note on lexical
  disambiguation"

James Cowie, "Deriving a vocabulary of
  part-names from a field guide text"
Willem Meijs, "Exploring lexical
  knowledge"