

# Briefly Noted

## Linguistic and Computational Techniques in Machine Translation System Design (second edition)

Peter Whitelock and Kieran Kilby  
(Sharp Laboratories of Europe; late of UMIST)

London: UCL Press (Studies in Computational Linguistics, edited by Harold Somers), 1995, xii+208 pp; hardbound, ISBN 1-85728-216-7, £30.00

This is a slightly enhanced second edition of the 1983 report 84/2 of the Centre for Computational Linguistics at UMIST in Manchester, England. As Harold Somers states in the preface to this edition, "it was the first attempt anyone had made to take a critical and objective look at the details of MT systems." This explains to a large extent the interest it may arouse and justifies to some extent its (re)publication within the Studies in Computational Linguistics series a decade after its original completion.

The book comprises two parts. Part I is a concise general discussion about machine translation (MT). It covers succinctly the linguistic and computational information necessary to be able to follow the second part of the book. Part II surveys, in great detail, and evaluates six MT systems: SYSTRAN, TAUM-Météo, TAUM Aviation, GETA Ariane-78, Yorick Wilks's Preferential Semantics (PS) system, and METAL. Neither of the authors were involved in the development of any of these software projects. They did, however, have an in-depth understanding of the intricacies of the systems, having collected numerous relevant reports and internal memos. Their description is given in a relatively clear, easy-to-follow manner without compromising on technical detail. Harold Somers has provided a number of additional footnotes trying to put (part of) the discussion into a mid-90's perspective.

The book is not, nor does it claim to be, either an introduction to MT or a general reference book on MT. In essence, it is a critical, detailed description of the six aforementioned MT systems. Somers claims in the preface that "this effort has not yet been duplicated." As humble as he may be, the facts are that by the time he wrote this preface (December 1993) Somers himself had already co-authored Hutchins and Somers (1992). The latter is a widely available introductory MT

textbook and reference book that contains, inter alia, four full chapters dedicated to critical, detailed descriptions of most of these MT systems (all excepting TAUM Aviation and Wilks's PS system). Although Hutchins and Somers may seem at times to exhibit more of a socio-historical interest than a technical one—in contrast to the book under review—nonetheless, they too go into substantial technical detail and give overall comparable descriptions. The effort, thus, *has* actually been duplicated to a significant extent, casting certain doubt on any serious need for this second (partially redundant) edition of Whitelock and Kilby's book.

To summarize, the book is interesting, well written, not too thick, and easily read. However, unless one is explicitly interested in detailed critical descriptions of TAUM Aviation or Wilks's PS system, then Hutchins and Somers (1992)—which covers most of that discussed by Whitelock and Kilby and much much more—should probably suffice. This renders the book under review somewhat dispensable.—*Daniel Radzinski, Lexical Technology Inc., Alameda, California.*

### Reference

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

## Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing

Stefan Wermter, Ellen Riloff, and Gabriele Scheler (editors)

(University of Hamburg; University of Utah; and Technical University of Munich)

Berlin: Springer (Lecture Notes in Artificial Intelligence 1040), 1996, ix+468 pp; paperbound, ISBN 3-540-60925-3, no price listed

This book proved itself, literally, to be a very effective survey of the recent work in a variety of approaches to learning natural language. I began reading this book while on a flight to attend the third International Colloquium on Grammatical Inference (ICGI-96), where I was invited to give a lecture on

grammatical inference of natural language. By the end of the book, and the flight, I had completely rethought the introduction to my talk.

The book is a collection of revised papers from a workshop at the 1995 International Joint Conference on Artificial Intelligence, with a very worthwhile introduction written by the members of the workshop organizing committee. As the title implies, the book is divided into three sections: connectionist networks and hybrid approaches, statistical approaches, and symbolic approaches to language learning.

Most of the selections in the first section are proof-of-concept papers, in contrast to the more results-oriented papers in the other two sections. The reader is left with the impression that while there is interesting work being done in applying neural networks and genetic algorithms to language learning, the approach is still in its infancy, and lags somewhat behind the statistical and symbolic movements. For example, the final paper in the section, "Learning language using genetic algorithms," by Smith and Witten, established the *robustness* of its proposed learning algorithm by trying to learn the rule  $S \rightarrow NP VP$  from a half-dozen five- and six-word sentences.

With this book, Wermter, Riloff, and Scheler have done a commendable job of providing the community with a good starting point for exploring these very different approaches to the problem of language learning.—David M. Magerman, *Renaissance Technologies Corp.*

### **Instrumentation: An Introduction for Students in the Speech and Hearing Sciences (second edition)**

T. Newell Decker  
(University of Nebraska—Lincoln)

Lawrence Erlbaum Publishers, 1996,  
xx+163 pp; paperbound, ISBN  
0-8058-2186-4, \$24.50

Although written primarily for students in audiology, speech pathology, and related areas, Decker's book will also be helpful to any linguist who wants to learn the practical aspects of electronics for collecting and analyzing speech data. The book assumes that

the reader has no scientific or technical background. The topics covered include the following: the basics of setting up and connecting equipment; microphones and other transducers; amplifiers, attenuators, mixers, and filters; digital signal processing; spectrum analysis; digital and analogue tape recording; and instrumentation for studying the physiology of speech.—G.H.

### **Language Industries Atlas (second edition)**

Josephine A. Edwards and  
A. Geoffrey Kingscott (editors)  
(Praetorius Ltd, Nottingham, UK)

Amsterdam: IOS Press, 1997, v+440 pp;  
hardbound, ISBN 90-5199-252-1, Dfl 180.00,  
\$100.00, £68.00, DM 160.00; distributed in  
Japan by Ohmsha Ltd, ISBN  
4-274-90080-0-C3000

This Eurocentric guide to the language industries and some of the basic concepts of language builds upon the 1994 edition, but the organization is now encyclopedic: one simple alphabetical list. The listings include major companies, organizations, technologies, and media. A typical entry is one or two paragraphs, with addresses and phone numbers where appropriate. The coverage may be inferred from the first few *M* entries:

*Maaleht* [Estonian newspaper]  
Macedonian  
Machine translation  
*Machine Translation* [journal]  
Machine Translation Summit  
Magyar Rádió  
Magyar Televizio  
Maison du Dictionnaire (La)  
Malaysian National Institute of Translation  
Maltese  
Manx  
Manx Language Unit  
Mechanical and Electrical Engineering Terminology Centre

Unusually for a reference book, the *Language Industries Atlas* is set in a large font with much white space. (For example, the list of standard abbreviations for languages takes six pages for a table that would have required little more than a single page in a

conventional layout.) The physical size of the book could have easily been halved, significantly reducing the price, with no loss of clarity.—G.H.

### Recent Advances in Parsing Technology

**Harry Bunt and Masaru Tomita (editors)**  
(Tilburg University and Keio University)

Dordrecht: Kluwer Academic Publishers  
(Text, Speech and Language Technology series, edited by Nancy Ide and Jean Véronis, number 1), 1996, xv+415 pp;  
hardbound, ISBN 0-7923-4152-X, Dfl 195.00,  
\$128.00, £86.00

The book contains revised versions of papers presented at two workshops on parsing technologies in 1993 and 1994. The contents of the volume are as follows:

- "Parsing Technologies, and Why We Need Them" by Harry Bunt
- "Fully Incremental Parsing" by Mats Wirén and Ralph Rönquist
- "Increasing the Applicability of LR Parsing" by Mark-Jan Nederhof and Janos Sarbo
- "Towards a Formal Understanding of the Determinism Hypothesis in D-Theory" by James Rogers and K. Vijay-Shanker
- "Varieties of Heuristics in Sentence Parsing" by Makoto Nagao
- "Parsing as Dynamic Interpretation of Feature Structures" by Harry Bunt and Ko van der Sloot

- "Proof Theory for HPSG Parsing" by Stephan Raaijmakers
- "Efficient Parsing of Compiled Typed Attribute-Value Logic Grammars" by Bob Carpenter and Gerald Penn
- "Predictive Head-Corner Chart Parsing" by Klaas Sikkel and Rieks op den Akker
- "GLR\* — An Efficient Noise-Skipping Parsing Algorithm for Context-Free Grammars" by Alon Lavie and Masaru Tomita
- "Evaluation of the Tagged Text Parser, A Preliminary Report" by Tomek Strzalkowski and Peter Scheyen
- "Learning to Parse with Transformations" by Eric Brill
- "Estimation of Verb Subcategorization Frame Frequencies based on Syntactic and Multi-dimensional Statistical Analysis" by Akira Ushioda, David Evans, Ted Gibson, and Alex Waibel
- "Monte Carlo Parsing" by Rens Bod
- "Stochastic Lexicalized Tree-Insertion Grammar" by Yves Schabes and Richard Waters
- "The Interplay of Syntactic and Semantic Node Labels in Parsing" by David McDonald
- "Integration of Morphological and Syntactic Analysis based on GLR Parsing" by Hozumi Tanaka, Takenobu Tokunaga, and Michio Aizawa
- "Structural Disambiguation in Japanese by Case Structure Evaluation based on Examples in a Case Frame Dictionary" by Sadao Kurohashi and Makoto Nagao
- "Flowgraph Parsing" by Rudi Lutz
- "Predictive Parsing for Unordered Relational Languages" by Kent Wittenburg