

Briefly Noted

Computers and Writing: State of the Art

Patrik O'Brian Holt and Noel Williams
(editors)

(Heriot-Watt University and Sheffield City Polytechnic)

Oxford: Intellect Books (Great Britain only)
and Dordrecht: Kluwer Academic
Publishers (outside Great Britain), 1992, xi
+ 387 pp.

Hardbound, ISBN 1-871516-20-X (Intellect)
and 0-7923-1858-7 (Kluwer), \$114.00, £67.00,
Dfl 195.00

Computers and Writing: Issues and Implementations

Mike Sharples (editor)

(University of Sussex)

Dordrecht: Kluwer Academic Publishers,
1992, 224 pp.

Hardbound, ISBN 0-7923-1966-4, \$82.00,
£44.00, Dfl 125.00

Also published as a special issue of
Instructional Science, 21(1-3), 1992.

Most of the computer-based writing tools presently in use are quite dumb, relying on simple techniques to achieve their goals; however, development of intelligent writing aids is an important area in computational linguistics. Two recent books address many of the issues involved in the development of better writing tools.

Computers and Writing: State of the Art (CWSA), and *Computers and Writing: Issues and Implementations* (CWII) contain papers from the Third International Conference on Computers and Writing and the Fourth Conference on Computers and the Writing Process, respectively. The second book contains selected papers, which were further developed by the authors for publication. It is not clear whether the papers in the first volume underwent the same selection process, but the time that elapsed between the conference and the publication of the book suggests that the papers might have been revised since their presentation at the conference.

The papers cover a wide range of subtopics; as Sharples observes in his introduction, "This volume is deliberately eclectic" (CWII, p. 3). Research interests of participants include cognitive science, computer science, education, engineering, linguistics,

and philosophy. The majority of the contributors are from the United Kingdom.

CWII contains an introduction by the editor and 15 papers; CWSA contains a preface by Holt and 24 papers. Neither book has a concluding chapter to tie the strands together in any way, so the reader may go away with a sense of having been left hanging at the end. Perhaps the absence of concluding chapters is due to the difficulty of drawing any general conclusions from such a wide spectrum of topics. They range from designing writing tools to assessing educational tools, from observing writing practices to writing interactive fiction. Thus, many of the chapters might not be of interest to any one reader, unless that reader wants to gain an overview of the issues that are currently being explored.

Most of the papers in the two books can be (very) loosely divided into three categories: the effects of the computer on writing practices; CAI for writing; and the development of computer writing tools. Neither of the books has a computational linguistics focus. About one-fifth of the papers in each describe the application of natural language processing techniques. For the most part, only current, reliable capabilities are applied to create working systems. I will briefly discuss a few of the papers that describe the implementation of NLP techniques, to give an idea of how they are being used.

Kempen and Vosse (CWSA) present a "language-sensitive" text editor for Dutch. Their system focuses on detecting real-word spelling errors that are homophonous to the target word, as well as flagging inconsistent spellings and other orthographical errors. The text undergoes both word-level and sentence-level analyses. At the word level, triphone analysis is performed to select all possible homophonous word alternatives. Next, a shift-reduce parser analyzes each sentence. The parser, which contains about 500 augmented phrase structure rules, ignores aspects of syntactic structure that are not relevant to orthography. Feature matrices on the nonterminal symbols detect feature violations. The evaluation of the system is only preliminary, but it appears to be a practical solution to the problem of discovering many real-word errors using current parsing capabilities.

Hoard, Wojcik, and Holzhauser (CWSA) report on a simplified English checker devel-

oped to aid manual writers at Boeing. A tokenizer, a lexicon, and a parser using a Generalized Phrase Structure Grammar are used to enforce vocabulary, grammar, and style restrictions that require only syntactic analysis. The authors intend to eventually add semantic and pragmatic modules to help resolve ambiguity and enforce meaning restrictions on allowable vocabulary. Their comparison of manual writers rewriting text into simplified English with or without the checker suggests that this difficult task will be made easier with this automated support. However, the examples of the error messages given for various sentences reveal that the suggested alternatives are not sufficient for users to easily translate "unsimplified" English to simplified English, since only word replacements are given when often a reformulation of the sentence is necessary. In any case, current parsers seem well suited to the limitations imposed by simplified grammars.

Forster and Van Nest (CWII) describe a prototype for a hypertext-based tool to aid in the on-line expansion of design briefs—short descriptions of artifacts, which are later elaborated upon to produce more detailed descriptions. Their system has various components aimed at avoiding the main difficulty associated with design brief expansion: linking a semantic entity in the design brief to requirement specifications. The most important component in this system is the chunker—a chart parser with a semantically augmented grammar—which is used to chunk briefs into smaller phrases, extracting the longest syntactically correct ones that are semantically relevant to the user. Only a shallow semantic understanding is achieved, to ensure that the speed of the parser is reasonable. The aim of the chunker is to present the designer with any relevant phrases and enough context from the design brief to be able to expand upon all requirement specifications without having to refer back to the original document. The authors claim that this method is superior to a simple keyword search with a window of words on either side because the chunker provides a syntactically complete context. However, whether there is actually a significant difference between these two approaches has not been tested.

One of the problems I found in perusing these volumes was the lack of overt organization of the papers. This seems to have

been intentional, and perhaps necessary, because of the variety of subjects, but I found the absence of chapters grouped into sections rather bewildering. There is a logical sequence from one paper to the next, but the topics are so diverse that predicting where a particular paper might be is almost impossible, unless one happens to be reading an adjacent one. Finding papers is particularly confusing in CWII, because of the absence of chapter numbering. Also, neither volume has an index to help alleviate the confusion; each paper does, however, have an abstract.

The editing of CWSA is poor. There are numerous typographical errors (mainly real-word errors), and the text of some of the non-native English writers contains many syntactically awkward and some incorrect constructions (see Andersen and Holmqvist's paper, for example). The book thus inadvertently illustrates the need for the development of good language-sensitive text editors. There are also many spacing errors, inconsistencies, and irregularities, particularly in several figures. Some of the papers in CWII would have been improved with some minor syntactic revisions (e.g., Andeweg, Hoekstra, de Jong, and Kunst), but this volume was generally free from typographical errors.

These books are obviously intended for a general audience. Although I have focused in this review on areas of interest to computational linguists, most of the issues discussed in these volumes will only whet the appetite of readers with specific interests. Computational linguists who want to develop writing tools for use in the near future may be inspired by the conservative but successful application of present NLP capabilities. Those who are seeking new research projects might also be inspired by the variety of ideas and the obvious need to take some of them several steps further to create truly intelligent tools for writers.—*Angela Glover, Ontario Institute for Studies in Education*

Multilingual Multimedia: Bridging the Language Barrier with Intelligent Systems

Masoud Yazdani (editor)
(University of Exeter)

Oxford: Intellect Books, 1993, x + 210 pp.
Paperbound, ISBN 1-871516-30-7, £14.95

Multilingual Multimedia contains a few possibly interesting ideas, but they are lost in the hype, the redundancy, the poor presentation, and the failure to distinguish speculation from completed research that characterize the book. Properly edited and presented, the material in this book might have made an interesting journal paper and a couple of good proposals to submit to a funding agency.

Most of *Multilingual Multimedia* is a report of a long-term project to build systems—LINGER (Language INdependent Grammatical Error Reporter) and its descendent eL (enhanced LINGER)—that would assist learners of a second language by detecting syntactic errors in their writing and helping the students to correct them. LINGER was first implemented for French, and, having been judged “general enough to be a good basis for other Romance Languages” (p. 60), it was then extended for “Italian, Spanish, German, and English” (*ibid.*). But despite the title of the book, the systems are not multilingual in the sense of dealing with more than one language at a time. Nor are they multimedia; that part is admitted to be still in the early stages of development (p. 1). And they are language-independent only in that language-dependent knowledge is confined to the lexicon and grammar modules, as one would expect in any parser. Although there is mention of the systems being tested with students, they apparently remain prototypes.

The story of LINGER and eL is presented as a 150-page sequence of chapters that appear to be papers and reports written as the project progressed, with no attempt to integrate them or to eliminate redundancies and out-of-date information. For example, the chapter by Yazdani and Uren, originally published in 1988, speaks of work then in progress and future plans for the system (p. 70). O’Brien’s chapter is written in the style of a project report by an undergraduate who has yet to master the basics of English punctuation. It consumes more than a

quarter of the book, giving equal emphasis to the important points (the system’s heuristics for correcting syntax errors), the conventional (morphological analysis), and the trivial (the necessity of the period at the end of Prolog clauses).

The one chapter in this sequence that is not in need of serious revision is Bolt’s, in which English-language LINGER is compared with six commercial grammar checkers on a corpus of errors said to be typical of language learners. Some of the test sentences are almost word salad; for example, *An young dog arent eats the smelly fish*. Not surprisingly, LINGER did better than the commercial programs, which are intended primarily for fluent or near-fluent users of English and are not so rigorous in trying to produce a full parse of each sentence. Bolt analyzes the architecture of the various programs, and concludes that it is inherent in the design of the commercial programs that they are unsuitable for use as an aid to language learners. LINGER is not without its problems either. Long sentences can be troublesome for LINGER (p. 145), and the corrections it suggests are not always those that assume the fewest errors; for example, *This books is very old* is corrected to *These books are very old* instead of *This book is very old* (pp. 150–151).

But throughout these chapters, it is never made clear what is supposed to be so special about LINGER. Parsing ill-formed input and diagnosing syntactic errors have, after all, been well-established research topics for more than a decade.¹ But none of the papers in this book, not even Bolt’s, makes any comparison with other research. In fact, none acknowledges that it exists.

The other four chapters of the book are merely research proposals. They are, at

1 See, for example, the special issue of the *AJCL* (1983) on ill-formed input. The well-known EPISTLE project on parse fitting and text critiquing (later named CRITIQUE) is described by Jensen, Heidorn, Miller, and Ravin (1983) and Ravin (1988). Catt and Hirst (1990) describe a system with pedagogical motivations similar to those of LINGER that includes new methods for syntactic error diagnosis and uses a grammar of the student’s native language to check for interference from that language when explaining errors—a method first applied by Schuster (1986). The same idea is mentioned on page 206 of the present book as a planned extension of eL.

least, better matched to the title *Multilingual Multimedia*. Pollard and Yazdani propose a HyperCard front-end to eL in which the student would be led through simple dialogs with text, pictures, and recorded speech in a choice of languages. Only a small demonstration prototype has been implemented. Yazdani and Pollard propose the creation of a multilingual, multimedia database of "learning materials," in a discussion that brings together pedagogical, political, and user-interface considerations. And in two chapters (the second of which repeats much of the material of the first), Mealing and Yazdani propose language-independent computer-mediated communication using icons instead of natural language. By returning a form that is printed in the book, one may obtain a demonstration disk with a HyperCard stack showing how a hotel booking might be negotiated in such a set-up.

The book is attractively designed and typeset (which is all too rare when camera-ready pages are produced by the authors), and except for O'Brien's chapter and the editor's preface, most of the text appears to have been copyedited. The appendix referred to on page 21 is missing, and no index is provided.—*Graeme Hirst, University of Toronto*

References

- AJCL (1983). Special issue on ill-formed input. *American Journal of Computational Linguistics*, 9(3-4), July-December 1983, 123-196.
- Catt, Mark, and Hirst, Graeme (1990). "An intelligent CALI system for grammatical error diagnosis." *Computer Assisted Language Learning*, 3, November 1990, 3-26.
- Jensen, Karen; Heidorn, George E.; Miller, Lance A.; and Ravin, Yael (1983). "Parse fitting and prose fixing: Getting a hold on ill-formedness." *AJCL* (1983), 147-160.
- Ravin, Yael (1988). "Grammar errors and style weaknesses in a text-critiquing system." *IEEE Transactions on Professional Communication*, 31(3), 108-115.
- Schuster, Ethel (1986). "The role of native grammars in correcting errors in second language learning." *Computational Intelligence* 2(2) (Special issue on AI approaches to education), May 1986, 93-98.

Íslensk Orðtíðbibók [Frequency Dictionary of Icelandic]

Jörgen Pind, Friðrik Magnússon, and Stefán Briem

(Institute of Lexicography, University of Iceland)

Reykjavík: 1991, lv + 1207 pp.

Hardbound, no ISBN (order directly from the Institute of Lexicography, University of Iceland, PO Box 7220, 127 Reykjavík, Iceland), \$95.00

The *Frequency Dictionary of Icelandic*, a project of the Institute of Lexicography of the University of Iceland, is based on an automatically tagged corpus of 100 contemporary 5,000-word texts selected from a variety of sources. The results are presented in a manner similar to Johansson and Hofland's analysis of the Lancaster-Oslo/Bergen corpus of English. For those who work with Icelandic but are not fully fluent in the language, an English translation of all the prefatory and introductory matter is available as the Institute's technical report number 4, which is included with the dictionary upon request.

Everything That Linguists Have Always Wanted to Know about Logic but Were Ashamed to Ask (Second Edition)

James D. McCawley

(University of Chicago)

Chicago: The University of Chicago Press, 1993, xxi + 633 pp.

Hardbound, ISBN 0-226-55610-7, \$100.00; paperbound, ISBN 0-226-55611-5, \$34.95

The new edition of McCawley's famous book is 25% larger than the first because "the range of things in logic that linguists in 1993 have always wanted to know is considerably larger than the range of things that linguists in 1981 had always wanted to know" (p. xi). New or expanded topics include the logic of conditional sentences, type theory in semantics, Gupta's "principles of identity," generalized quantifiers, λ -calculus, and conventional implicature. In addition, much of the earlier material has been rewritten and new exercises added.

CHILDES/BIB: An Annotated Bibliography of Child Language and Language Disorders. 1994 Supplement

Roy Higginson and Brian MacWhinney
(University of California, San Diego and Carnegie Mellon University)

Hillsdale, NJ: Lawrence Erlbaum Associates, 1994, vi + 693 pp.
Hardbound, ISBN 0-8058-1478-7, \$125.00

The first volume of this bibliography on child language and language disorders was published in 1991. This new supplement covers North American research since late 1990 and international research published outside North America since 1970.

Conceptual Structures: Current Research and Practice

Timothy E. Nagle, Janice A. Nagle, Laurie L. Gerholz, and Peter W. Eklund (editors)
(Unisys Corporation; School District 833, St Paul Park, MN; Unisys Corporation; and University of Adelaide)

Chichester: Ellis Horwood (Ellis Horwood Series in Workshops), 1992, xv + 644 pp.
Hardbound, ISBN 0-13-175878-0, no price listed

This book contains 29 papers from the Fifth Annual Workshop on Conceptual Structures (1990), nine of which concern natural language and text retrieval:

"Lexical choice as pattern matching" by Jean-François Nogier and Michael Zock,
"Analysis and generation of Italian sen-

tences" by F. Antonacci, Maria Teresa Paziienza, M. Russo, and Paola Velardi,
"Representing word senses for semantic analysis" by Maria Teresa Paziienza and Paola Velardi,
"Conceptual parsing: A syntax-directed joining algorithm for natural language understanding" by Stéphane Bornerand and Gérard Sabah,
"Text generation in expert critiquing systems using rhetorical structure theory" by Jonni Harrius,
"Generating language from conceptual dependency graphs" by Afke van Rijn,
"The dynamic type hierarchy theory of metaphor" by Eileen C. Way,
"Conceptual graphs as a framework for text retrieval" by Sung H. Myaeng, and
"Conceptual graph information retrieval using linear resolution, generalization and graph splitting" by Jean Fargues.

The New Hacker's Dictionary (Second Edition)

Eric S. Raymond (compiler)

Cambridge, MA: The MIT Press, 1994, xxi + 505 pp.
Paperbound, ISBN 0-262-68079-3, \$14.95

Yu-Shiang Whole Fish n. obs. The character gamma (extended SAIL ASCII 00010001), which with a loop in its tail looks like a little fish swimming down the page. . . . Usage: primarily by people on the MIT LISP Machine.

zen vt. To figure something out by meditation or by a sudden flash of enlightenment.
zeroth adj. First.