

lents in classical or nonclassical logical languages, and for which a logical translation into terms of logical constants is impossible (“*il existe un grand nombre de connecteurs en langue naturelle n’ayant pas de correspondants dans les langages logiques classiques ou non classiques et pour lesquels la traduction logique en termes de constante logique n’est pas possible*”). With these words Moeschler subscribes to the implicitness principle (Schubert 1988b), which basically says that an artificial representation has an inherently insufficient expressive power for rendering the full and unrestricted content of a text in a human language. Some of Moeschler’s examples illustrate very well how the content of connector constructions is impoverished when rendered in predicate logic.

It would be a promising continuation of the work begun by Moeschler if his connector analyses were applied cross-linguistically and, in particular, in translation studies. The implicitness principle shows that expanding the scope of work in this way by no means requires any logical representation to be chosen as the main object of research rather than the words themselves. On the contrary, a cross-linguistic investigation of translation equivalences can, in my view, profit from an approach like Moeschler’s.

Part 1 of the book analyzes the logical content of connectors, Part 2 discusses their function as markers of pertinence and, indirectly, of coherence, and Part 3 suggests a tree-structured discourse representation. This representation might be seen as one of many (for an overview of text models see Papegaij and Schubert 1988: 13–14), if it were not for the specific view on connectors that underlies Moeschler’s model. With the exception of a brief reference to scripts, the book at first sight does not seem to have any direct link to computational linguistics. Nevertheless, I am convinced that a study that approaches language at the grammatical and pragmatic level and leads toward a linguistically motivated formalization is worth the attention of computational linguists. I feel that materials of this kind are needed on a much larger scale than is currently available, to let computational linguists’ skills in formalization and implementation be applied on a pertinent linguistic basis.

REFERENCES

- De Beaugrande, Robert-Alain and Dressler, Wolfgang Ulrich. 1981 *Einführung in die Textlinguistik*. Tübingen: Niemeyer.
- Papegaij, Bart and Schubert, Klaus. 1988 *Text Coherence in Translation*. Dordrecht and Providence: Foris.
- Schubert, Klaus. 1988a *Ausdruckskraft und Regelmäßigkeit*. *Language Problems and Language Planning*, 12, 130–147.
- Schubert, Klaus. 1988b *Implicitness as a Guiding Principle in Machine Translation*. *Proceedings of the 12th International Conference on Computational Linguistics (COLING-88)*. Budapest, 599–601.

Klaus Schubert holds a doctorate in General, Scandinavian, and Slavic Linguistics from the University of Kiel (Federal Republic of Germany). In 1985 he joined the DLT machine translation project at the BSO software house in Utrecht. Since 1989 he has been director of BSO/Research and leader of the DLT project.

Schubert’s address is: BSO/Research, Postbus 8348, NL-3503 RH Utrecht, The Netherlands. E-mail: schubert@dlt1.uucp

COMPUTATIONAL LINGUISTICS: AN INTERNATIONAL HANDBOOK ON COMPUTER ORIENTED LANGUAGE RESEARCH AND APPLICATIONS / COMPUTERLINGUISTIK: EIN INTERNATIONALES HANDBUCH ZUR COMPUTERGESTÜTZTEN SPRACHFORSCHUNG UND IHRER ANWENDUNGEN

István S. Bátori, Winfried Lenders, and Wolfgang Putschke (eds.)

Berlin: Walter de Gruyter, 1989, xxxiii + 933 pp. (Handbooks of Linguistics and Communication Science 4 / Handbücher zur Sprach- und Kommunikations-wissenschaft) Hardbound, ISBN 3-11-009792-3, DM 728.-

Reviewed by
Graeme Hirst
University of Toronto

First, the form of the book: The *Computational Linguistics Handbook* is very large. It contains 65 papers, organized in 12 sections, covering all aspects of computational linguistics in the widest sense of the term. It has nearly a thousand pages, most in two-column format, and weighs 2.2 kg. It’s probably the first book in computational linguistics to include fold-out maps.

The book is in English and German: front matter is in both languages, and articles are in one or the other (about half in each language). One would have hoped that each article would be accompanied by at least a translation of the title and a short abstract in the other language, so that the unilingual user could judge the relevance of the article and possibly seek assistance with translation. But, unfortunately, only titles are translated, and only in the table of contents, not in the article itself. This makes browsing difficult for users who do not have some facility in both languages—and one of the values of the book is that it invites browsing.

The book is properly typeset throughout—there’s no author’s rough camera-ready copy here—which accounts at least in part for its high price. Nevertheless, typos (or copyediting infelicities) turn up a little more often than one would have hoped. And on a couple of occasions, the running heads are completely out of sync with the text.

Now the content: The goal of the *Computational Linguistics Handbook* (pp. xiv–xv) is to describe comprehensively the current state of research in the field, to survey the literature, to locate the field itself with respect to related disciplines and show its application in those fields, and to suggest how the field may be further developed. That’s a big goal for a big book.

In the space for this review, I cannot even list each article

of the handbook, let alone discuss each. But a work like this succeeds or fails not so much with the individual articles as with the whole that they create, and I shall concentrate on that. I will follow the structure of the book, which is based on the editors' view of the structure of the field.

The first three parts of the book (126 pages) are background. Part I, with articles by Bátor, Endres-Niggemeyer, and Hajičová, covers CL as an enterprise: the history and organization of the field, and its connections to other disciplines. Included is a long listing of many CL research groups around the world, giving representative publications of each. Fold-out maps mark the location of each group (not always with geographical precision), using colored symbols to indicate the specialties of each. There is a special survey of CL research in Eastern Europe. Parts II and III, with articles by Martin, Krallman, Klenk, Bertsch, Karlgren, Köhler and Gabriel Altmann, and Krause, describe the foundations of the content of CL and its methodology: what CL aims to do, what the basic problems studied are, the tension between research and application, the role of quantitative methods.

The next five parts cover basic research, and are divided into three parts (IV, V, and VI) on issues in the static description of language and two parts (VII and VIII) on issues in processing, entitled 'computer simulation of language.'

The parts on description (133 pages) cover all levels from phonological segmentation of speech up to Schankian scripts. Several articles discuss the value of corpus-based research, and there is a survey of corpora that are available for research. Formalisms and theories of morphology, syntax (ATNs, DCGs, unification grammars), and semantics (semantic markers, conceptual dependency, etc.) are included. Two articles (those of Part VI) emphasize the testing and validation of theoretical descriptive models by computer implementation of them. The articles are by Sgall, Bergenholtz and Mugdan, Allen, Lenders, Piirainen, Wothke, Schraeder and Willée, Ramsay, Karlsson, Friedman, and Eikmeyer, Kindermann, and Petöfi.

Part VII (51 pages), with articles by Lenders, Gerry Altmann, Öim, Berry-Rogghe, and Metzinger and Görz, is the first on processing. It discusses foundational issues, the relationship between CL and the neighboring disciplines of artificial intelligence and psycholinguistics, and the role of knowledge representation formalisms in general and semantic networks in particular. Part VIII (170 pages) then presents the components of natural language understanding and generation systems; the articles are by Bátor, Wettler, Ingria, Hellwig, Rollinger, King, Lutz, and Kempen. There is first an overview of the various architectures that have been proposed for NLU systems, and of the many varieties and flavors of knowledge representation. Then individual system components each get an article: lexicons, parsers (in very considerable detail), semantic interpreters, and inference engines. Generation is covered in a single short article.

The next three parts, IX to XI, cover applications of CL, and it is here that we see how wide the field is, or may be

construed to be. Antonio Zampolli has observed (in an address to the International Conference on Computers and the Humanities, June 1989) that Italian has only a single term, *linguistica computazionale*, for the enterprises that in English are divided into "computational linguistics" and "linguistic computing," and that research in Italy has benefited from the unity of *linguistica computazionale*. Likewise, in this handbook, the applications presented serve to emphasize the continuity between what anglophones regard as separate fields.

Part IX contains a single short article by Bátor introducing applications of CL. The applications presented are divided into those that involve the description of language (Part X), and those involving any aspect of human-machine communication (Part XI). The first division (96 pages) covers lexicography (including indexes, concordances, dictionaries, and lexical databases), grammaticography, dialectology (more fold-out maps!), reconstruction of texts, authorship attribution, the analysis of style, historical linguistics, and content analysis. The articles are by Jones and Sondrup, Calzolari, Wegea and Berg, Wickmann, Najock, Allen, Hewson, Mohler, and Händler, Hummel and Putschke. The second division (166 pages) covers speech recognition and synthesis, machine and machine-aided translation (six articles, including surveys of current systems), information retrieval (two articles, one of them a survey of current systems), question-answering systems (two articles), office automation, and computer-assisted instruction. The authors are Ney, Müller, Melby, Slocum, Fábrič, Tsujii, Boitet, Marchuk, Kuhlen, Panyr and Zimmermann, Habel and Pribbenow, Lesch, Ballard, and Zettersten.

Part XII (47 pages) is about the 'operating environment' of CL, which is taken to include hardware and software (programming languages, graphics packages), data interchange standards, and the like. The articles are by Hockey, Drewek, Messerschmidt and Hotz, Seppänen, and Händler.

A work of this kind needs a first-rate bibliography and subject index. The bibliography here is 101 pages long, and contains about 3,700 entries. But the subject index is not so successful. It contains English and German terms all mixed up together, most without cross-reference; the user must look for both—for example, "question-answering-system" and "Frage/Antwort-Systeme." Incidentally, the latter entry, but not the former, includes a listing of all the question-answering systems mentioned in either language in the text. Cognate terms are conflated under the German; for example, there is no entry at "grammar," but a long one a few lines later at "Grammatik / grammar." In the case of terms that are relatively distant alphabetically, such as "Regel / rule," the user who knows no German is rather at a disadvantage. The index terms are those "suggested by the authors," rather than by an indexer who may have had a clearer picture of the whole work and the access points that a reader may need, and who may have imposed a greater uniformity of terminology. For example, preference semantics is discussed in (at least) two articles; one is

indexed under "Präferenzsemantik," the other under "preferences."

A work as big as this is sure to contain something to offend everyone. Some of the articles, rather than being well-fitting pieces in structure, are perhaps a little too idiosyncratic; for example, Allen's lengthy complaints about the paucity of computer-assisted stylistic studies of Spanish text. And some authors are more adept than others at making their topic comprehensible to the newcomer; the same article by Allen is particularly reader-friendly. But despite the handbook's size, what I noticed most were the omissions. The short article on computer-assisted language teaching does little justice to current research into the application of sophisticated CL methods to the problem. I could find nothing on dealing with ill-formed input (or did the index let me down?). Transformational grammar is mentioned a number of times, but government-binding theory is not. Research in discourse structure is hardly mentioned, except, unexpectedly, in the article on language generation. No form of the term "anaphora" appears in the index (though there is at least a passing mention of the problem (p. 270) in Lenders's introduction to the sections on processing). But such complaints should not be allowed to obscure the wealth that can be found in this handbook. I just wish that it were a little easier to find what one is looking for.

Graeme Hirst is the book review editor of *Computational Linguistics*. Hirst's address is: Department of Computer Science, University of Toronto, Toronto, Canada M5S 1A4. E-mail: gh@ai.toronto.edu

BRIEFLY NOTED

ALTERNATIVE CONCEPTIONS OF PHRASE STRUCTURE

Mark R. Baltin and Anthony S. Kroch (eds.)
(New York University and University of Pennsylvania)

Chicago: University of Chicago Press, 1989, xi + 315 pp.
Hardbound, ISBN 0-226-03641-3, \$60.00; Softbound, ISBN
0-226-03642-1, \$19.95

An essential assumption of every syntactic theory is that the sentences of natural language are internally structured. In the early years of generative grammar it was generally believed that the appropriate mechanism for generating syntactic structures was a grammar of rewriting rules that operated free of context. Recently, however, this belief and others concerning the nature of phrase structure have increasingly come into question.

The twelve essays in this volume grew out of a conference convened to discuss the challenge to the classical formulation of phrase structure and the alternative conceptions proposed to replace it. Each of the articles approaches this issue from the perspective of a different linguistic framework, such as categorial grammar, government-binding theory, head-driven phrase structure grammar, and tree-adjoining grammar. Evidence from a number of languages and subdomains of grammar are brought to bear on a variety of questions. What are the primitives out of which structured representations are constructed?

What is the formal device that generates the infinite set of

representations for the infinite set of sentences of each natural language? What is the relationship between lexical selection and structure generation? The contributors also consider the possibility of reassigning the work done by rewriting rules to other formal devices, the nature of selection, the character of unbounded dependencies, and the treatment of word order variation in so-called free word-order languages.—*From the publisher's announcement*

PROCEEDINGS, SPEECH AND NATURAL LANGUAGE WORKSHOP

Defense Advanced Research Projects Agency

San Mateo, CA: Morgan Kaufmann Publishers, February 1989,
vii + 295 pp.
Softbound, ISBN 1-55860-073-6, \$35.00

The workshop reported here brought together 17 DARPA contractors in spoken language recognition and natural language processing research. Included are a summary of the proceedings and technical papers from most of the groups.

NATURAL LANGUAGE PROCESSING TECHNOLOGIES IN ARTIFICIAL INTELLIGENCE: THE SCIENCE AND INDUSTRY PERSPECTIVE

Klaus K. Obermeier
(Battelle Laboratories)

Chichester, England: Ellis Horwood, 1989, 263 pp.
(Ellis Horwood Series in Artificial Intelligence)
Hardbound, ISBN 0-7458-0562-0 and 0-470-21528-3, \$39.95

This book reviews natural language processing technology viewpoints from both the academic and the business perspective. It is not a highly technical or academic book per se, but rather, it reviews the technology as a whole for the nonexpert in natural language. The book presents the state of the art, forecasts of the technology, and the outlook from a commercialization standpoint. The reader who would benefit most is the technical manager or professional contemplating the commercialization of an NLP product. If the book is not mistaken as a "how to" textbook, it may prove useful and satisfying; similar industry reports and forecasts from private sources may cost hundreds of dollars.

A broad range of topics is covered. The book begins with an overview of NLP approaches, formalisms, and methods. This is followed by a detailed coverage of applications from natural language interfaces and machine translation to text processing, generation, writing aids, and speech. For each, the approaches, state of the art, existing products, and prognosis for future products are presented. The book ends with a section on current issues, the business picture, and the "science" of NLP. A major thesis of the book is that there is a marked difference between business people and academics in the field; breakthroughs in the science are needed and they should be driven by practical market needs. For example, the metric of success in NLP should be productivity improvement, not linguistic perfection.

The overall range of topics and information makes the book useful from a practical standpoint. However, there are some problems with the writing style and content in spots. The author occasionally writes paragraphs that are incoherent—that is, they wander into a number of different unrelated topics. Further, some of the information was not carefully researched. Two examples: On page 109, the author implies that the Unix Consultant is a system that translates English commands to Unix, when it is actually an advisory system about Unix. On page 119, a natural