

expression or syntactic rule in one there is a corresponding expression or syntactic rule in the other with the same meaning. Rosetta differs from the strict Montague model in making no use of intensional logic or of categorial grammar, but it adheres to the compositionality principle (in relating syntactic rules and semantic structures). The paper includes many formal proofs and only trivial sentence examples—which might well not impress those skeptical of the value of a project whose influence on MT theorists has been considerable in recent years. It is followed by a description of another experimental system (albeit less radically innovative in linguistic conception), an account by Margaret King and Sergei Perschke of the Eurotra project of the European Communities. After outlining the historical background (the need for a decentralized multilingual project capable of stimulating MT and computational linguistic research in each of the participating countries), the basic features are described: a linguistics-based transfer model, modular and robust, no interactive facilities, multilevel tree-structure interfaces, a controlled production system operating on series of well-defined grammars, etc. Eurotra is admitted to be relatively traditional in its linguistics, but it is claimed to be more advanced in its computational design and to have been successful in promoting the study of languages not previously the subject of MT or detailed linguistic research.

This book represents a primary source of information for some of the most significant and influential MT projects in the last decade. It is greatly enhanced by a substantial bibliography, which includes not only all references by the contributors, but also many other items covering the historical development of MT, the current (1984) state of research, and related topics in computational linguistics and artificial intelligence. The publication delay has not diminished the value of the contributions: many papers provide the most detailed and comprehensive accounts of individual MT systems. Only in the cases of Rosetta, GETA, and Eurotra can it be said that there have been substantial changes in the years since the Lugano tutorial. There are some omissions in the historical surveys as already noted, and readers should also be made aware that a number of systems current at the time are not mentioned at all, e.g., the DLT system in Utrecht, the PAHO system in Washington, many of the largely AI-inspired projects in the United States, the systems in the Soviet Union, and the activity in Japan (e.g., the influential Kyoto research). These are minor points. This is a publication that will remain for many years an invaluable source on MT research and it deserves to find a place on the shelves of anyone seriously interested in computational linguistics and in machine translation.

*John Hutchins* is the author of *Machine translation: Past, present, and future* (published by Ellis Horwood). His address is The Library, University of East Anglia, University Plain, Norwich NR4 7TJ, England. E-mail: L101@uea.cpc865

## ADVANCES IN NATURAL LANGUAGE GENERATION: AN INTERDISCIPLINARY PERSPECTIVE

Michael Zock and Gérard Sabah (eds.)

(LIMSI, Orsay)

London: Pinter Publishers and Norwood, NJ: Ablex Publishing, 1987, 2 vols., xix+200 pp., xviii+176 pp.

(Communication in Artificial Intelligence Series)

ISBN 0-86187-965-1 and -995-5, £27.50 per volume (hb); ISBN 0-89391-527-0 and -537-8, \$45.00 per volume (corporate), \$27.00 per volume (personal)

*Reviewed by*

*Marie Meteer*

*BBN Systems and Technologies Corporation*

*Advances in Natural Language Generation* is the proceedings of the European Workshop on Natural Language Generation held in January 1988 at the Abbey de Royaumont, France. As such, it gives a fairly broad snapshot of the work in this area going on in Europe (with a few notable omissions, such as Wahlster, Reithinger, and Danlos). The introduction by editors Zock and Sabah, and the foreword by David McDonald emphasize the youth of natural language generation (NLG) as a field and the importance of workshops such as this one in bringing together researchers from diverse areas, such as psychology, linguistics, and artificial intelligence, to contribute to this growing field.

Perhaps due to the youth of the field or perhaps due to a lack of rigor on the part of the workshop organizers and editors, the book lacks unity. The papers themselves vary a great deal in quality and in audience, some presenting an introduction to some aspect of the field and others assuming the reader knows the particular details of a linguistic framework. Some address themselves directly to the problems of NLG, while others report on just their own work and leave the connections to NLG to the reader.

The first paper in the book, "Language Generation and Explanation" by McKeown and Swartout, is aimed at a wide audience and presents issues and approaches rather than particular advances in the field. It is not a paper from the workshop itself, but rather a reprint from the *Annual Review of Computer Science*, and as such is a very good general overview of previous work, particularly in text planning and explanation. Unfortunately, it does not provide adequate background for the papers in this collection; in particular, it does not supply any introduction to the different kinds of grammars in use, such as LFG, SFG, and FUG.

As the title of the book suggests, many of the papers are reports on work that has in some way furthered the state of the art in NLG. These papers are aimed at those who already know the field, at least to some degree, and who know some of the background behind the problems being addressed. The best of the papers were those that

tackled theoretically interesting problems, defined them narrowly enough to be covered adequately in a short paper, grounded their work in implemented systems performing a specified generation task, and provided comparisons with other work in the field. Unfortunately, few of the papers fit these qualifications. Three in particular stood out: Novak's "Generating Referring Expressions in a Dynamic Environment", Mellish's "Natural Language Generation from Plans", and Houghton and Pearson's "The Production of Spoken Dialogue".

Novak's paper addresses the problem of selecting the information required to uniquely identify an entity in the world and expressing it cohesively in a descriptive paragraph. The structure of his domain, the description of objects moving in a street scene, allows interesting aspects of these problems to be worked on. For example, descriptions of entities can include the events they have participated in as well as their attributes.

Mellish also works within a particular domain (planning) and a particular generation task (producing instructions to carry out those plans). His work focuses on defining an architecture for generation that uses intermediate levels of representation between the stages of processing. In this work he posits a process before generation begins, called "optimization", which reorganizes the random structures of the application program into something usable by the generator. His use of predefined "algebraic simplifications" to optimize intermediate levels of representation rather than searching through all possible solutions to find an optimal one makes the task tractable.

Houghton and Pearson address the problems of producing spoken language. They work within the context of simulated robots cooperating to complete a task, which provides motivations for situations and dialogs in which to study the subtle distinctions expressed by intonation.

The papers in the book are divided into four categories: Linguistic Approaches, Implementation Issues, Psychological Issues, and Educational Applications. Nearly half of the papers are in the Implementation Issues section. They range from complete implementations, such as Roesner's detailed discussion of the four application areas that the SEMSYN project has generated text for, to Kwee's anecdotes about extending a functional unification grammar, complete with all the incorrect paths he took by not adequately analyzing the phenomena before implementing.

Many of the papers in this section went beyond implementation issues to more theoretical issues, such as Novak, Mellish, and Houghton and Pearson, to the point where the section would have benefited from being subdivided along those lines. Other papers didn't address implementation at all, but focused on an approach to specific problems in NLG. In particular, Dale looks at how the structure of discourse affects choosing

an appropriate referring expression, critiquing the work of Grosz and Sidner.

The section on psycholinguistic issues exemplifies a major contribution of this kind of workshop in bringing together psycholinguists with AI researchers. Two papers are particularly noteworthy. Harley, who is known for work in speech errors, presents a two-phase model of language generation: 'automatic' and 'executive' processing. Schriefers and Pechmann report on a study of incremental text production and its effect on referential noun phrases. The only flaw of these two papers is that they are too short. One is left wanting greater detail about both the studies themselves and their significance for the generation process overall.

The most disappointing section was that on linguistic approaches. Many of these papers neither further the field nor give the reader an overview of a particular area. They focus on the competence of a grammar with no consideration of how to choose between various forms. As Zock and Sabah begin the introduction to the book, "text production is decision making under specific . . . constraints". Papers such as Block's ("Can a 'Parsing Grammar' be used for Natural Language Generation? The Negative Example of LFG"), Horacek's ("The Application of Unification for Syntactic Generation in German"), Bunt's ("A Phrase Structure with Discontinuous Grammar"), and Parisi and Giorgi's ("A Lexically Distributed Work Ordering Component") all focus on formalisms for competence grammars without considering how effective they are for the major process in generation: decision making. Furthermore, they are rather narrow in scope, not recognizing the work of others or telling the reader how their work fits into the larger picture. Of the grammar papers, Fawcett's "Language Generation as Choice in Social Interaction" addresses the problems of NLG most directly. He presents a high level comparison between systemic functional linguistics and what he calls "neo-Chomskyan" approaches, pointing out the need to consider all aspects of language and its relation to knowledge, rather than concentrating on structure alone.

Also in this section, Dik ("Concerning the Logical Component of a Natural Language Generator") addresses the problem of choice more directly and goes beyond discussing merely syntax. In fact, he includes in the generator work that most generation researchers would assign to the application program, for example, inferring the answer to a question from other information in the knowledge base. He also makes the assumption that a single representation throughout all the generation process (i.e., linguistic realization, knowledge representation, and the logical component that does the inferencing) "would, in fact, be ideal", without addressing the fact that other researchers have argued for more heterogeneous architectures (e.g., Mellish in the same volume, and McDonald, Meteer, and Pustejovsky 1987).

The final section is two papers on how generators can

be used in educational applications. Bakker, van der Korst, and van Schaaik put a new twist on using a generator to implement and test a linguistic theory: they propose using such a program to teach a particular theory (in this case lexical functional grammar) in a computational linguistics course.

An odd fact about the book is that it is published in two volumes, each relatively short (under 200 pages). The topics are split between the volumes, with half the papers from each category in each book. I could find no particular motivation for finding a paper in one volume or the other. The foreword, introduction, and contents are repeated in both books, and each has an index, which unfortunately is split between the volumes, so you need to look up topics twice to find all the papers that address them. Overall, the division into volumes is more annoying than useful (I inevitably found myself on the right page of the wrong volume when looking for a particular paper), and since there is no organization to the division, it is unlikely that someone would be interested in one volume and not the other.

Despite these complaints with the presentation and some of the papers, the book is useful for researchers in NLG and related fields to see the range of work going on in Europe. While the papers themselves are short, their bibliographies can lead the interested reader to a wealth of related material. Furthermore, workshops such as the one motivating this book are important and their results need to be published. In a young field such as NLG the focus of the workshops needs to be broad so that it can define what the major issues are. Collections such as this one and Kempen (1987) help other researchers see how their work interacts with the field as a whole. As the field matures, however, these workshops need to become more focused. As in the AAAI-88 workshop (Hovy, McDonald, and Young 1988), the organizers and the editors need to formulate questions and encourage researchers to address them directly. This will provide more unity in the resulting collections, allowing a reader to get a complete picture of issues rather than isolated snapshots of individuals' work.

#### REFERENCES

- Hovy, Eduard; McDonald, David D.; Young, Sheryl 1988 *Proceedings of the AAAI Workshop on Text Planning and Realization* (AAAI-88), St. Paul, MN.
- Kempen, Gerard 1987 *Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics*, Martinus Nijhoff, Dordrecht, The Netherlands.
- McDonald, David D.; Meteer, Marie W.; and Pustejovsky, James D. 1987 Factors Contributing to Efficiency in Natural Language Generation. In: Kempen (1987); 152–182.

*Marie Meteer* is an associate scientist working on natural language generation in the AI Department of BBN Systems and Technologies Corporation. She is the primary developer of the SPOKESMAN generation system, which combines a newly designed text planner and the linguistic realization component Mumble-86. She is also finishing her Ph.D. at the

University of Massachusetts at Amherst on decision making and revising in natural language generation. Meteer's address is: BBN Systems and Technologies Corporation, 10 Moulton Street, Cambridge, MA 02138. E-mail: mmeteer@bbn.com

---

#### NATURAL LANGUAGE PARSING SYSTEMS

Leonard Bolc (ed.)  
(Polish Academy of Sciences)

Berlin: Springer-Verlag, 1987, xviii + 367 pp.  
(Symbolic Computation and Artificial Intelligence Series)

ISBN 3-540-17537-7 and 0-387-17537-7 (hb)

*Reviewed by*  
*Petr Sgall*  
*Charles University*

This volume, containing nine systematic studies on several of the most advanced parsers, fulfills very well the editor's aim characterized in his short preface, namely to present detailed accounts of research by outstanding specialists. The need for a complex overview of the present state of research in the field is mentioned here as a point for subsequent work.

J.G. Carbonell and P.J. Hayes point out in their contribution (pp. 1–32) how a multi-strategy approach (dynamically changing the procedures according to kinds of constructions being parsed) together with fail-soft heuristics and with user interaction can be used for an efficient treatment of semantic and structural ambiguity. Two implemented experimental parsers integrating several parsing strategies are characterized: CASPAR, combining a case-oriented strategy (based on domain semantics) with linear pattern matching, and thus obtaining the flexibility and robustness needed to handle several kinds of deviations from grammaticality; and DYPAR, including a context-free grammar (with domain information grouped into hierarchical semantic categories), a pattern matching component, and a mapping into a canonical form (accounting for semantic equivalence). Furthermore, an integration of the strategies involved is discussed, yielding a multi-strategy algorithm that uses top-down case-frame expectations to constrain bottom-up pattern matching. ATNs with logical variables and unification are compared with definite clause grammars and used as the basic ingredients of the system described by T.W. Finin and M. Stone Palmer (pp. 33–48). A similarity between the applied framework and certain aspects of lexical functional grammar are pointed out.

The third contribution, by J.G. Neal and S.C. Shapiro (pp. 49–92), stresses the role of natural language understanding, pointing out that natural language serves as its own metalanguage. A system understanding sentences about the use of language is described, which includes a kernel language (expressing the knowledge necessary for a poor language user to be able to be