

Briefly Noted

Generating Natural Language Descriptions with Integrated Text and Examples

Vibhu O. Mittal

(Just Research and Carnegie Mellon University)

Mahwah, NJ: Lawrence Erlbaum Associates, 1999, xxv+260 pp; hardbound, ISBN 0-8058-2414-6, \$45.00; paperbound, ISBN 0-8058-2415-4, \$22.50

Examples play a crucial role in instruction manuals, textbooks, and documentation. If they are carefully chosen they can show clearly how a new facility or language can be used, while also supporting conceptual understanding. To be effective, examples should be linked with any textual explanation of a topic, complementing that description.

Language generation techniques have been used in several prototype documentation systems—e.g., Reiter, Mellish, and Levine (1995). Documentation can be generated on-the-fly, tailored to characteristics of the particular user, using a knowledge base that can (sometimes) be automatically updated as a system specification is revised. This provides the possibility of improved system maintainability and effectiveness. If such systems are to begin to challenge conventional methods for documentation, then it is essential that they provide integrated and well-chosen examples.

This book looks at how examples can be selected within a system to generate natural language descriptions. This builds on the work of the Explainable Expert Systems project (Swartout, Paris, and Moore 1992) and many of the examples illustrate a plan language used in that project. The book is based on the author's Ph.D. thesis and includes an interesting review of work on the use of examples, taken from several disciplines, some conclusions and illustrations from a corpus analysis of instructional texts, and a detailed description of an implemented system.

The book highlights many of the issues that need to be addressed when generating examples that are based on some underlying specification. A particular contribution is highlighting that the use (and categoriza-

tion) of examples should depend on any surrounding textual explanation.

Given the range of interesting issues discussed, the actual system description is a little disappointing. For example, the rather general issue of how the presentation of examples depends on text type is reduced to a simple expert/novice reader distinction, and the dependence of examples on the type of knowledge being conveyed is not addressed. The system, however, suggests a general direction and approach for integrated example generation, and highlights the issues and problems.

One problem that is highlighted by the book is the difficulty in generating effective examples given limitations in the underlying knowledge representation. The system described uses representations of concepts that are purely syntactic (simply the BNF definition of allowed constructs). As a result, some of the examples are a little odd. For example, a generated example of the use of the LISP function `let` is:

```
(LET ((ORANGES FISHES))
      MEN)
```

While syntactically acceptable, it is certainly rather odd! This highlights a general difficulty in generating effective descriptions—they can only be as good as your underlying representation allows.

Overall I found the examples and discussion in the book more interesting than the details of the actual system, which were occasionally obscured by the effort needed to follow some of the underlying representations. It is an interesting book to read, and I would recommend it both to those working on text planning and those with an interest in the use of examples.—*Alison Cawsey, Heriot-Watt University*

References

- Reiter, Ehud, Christopher Mellish, and John Levine. 1995. Automatic generation of technical documentation. *Applied Artificial Intelligence*, 9(3):259–287.
- Swartout, William R., Cécile L. Paris, and Johanna D. Moore. 1992. Design for explainable expert systems. *IEEE Expert*, 6(3):58–64.

Corpus Linguistics: Investigating Language Structure and Use

Douglas Biber, Susan Conrad, and Randi Reppen

(Northern Arizona University, Iowa State University, and Northern Arizona University)

Cambridge University Press (Cambridge approaches to linguistics, edited by Jean Aitchison), 1998, x+300 pp; hardbound, ISBN 0-521-49622-5, \$59.95; paperbound, ISBN 0-521-49957-7, \$19.95

This book provides fairly detailed analyses of two research questions in corpus linguistics. The first half discusses how a concordance, simple counting programs, and a corpus can be used to identify differences in the rate of occurrence of a linguistic feature across different registers or conditioned on another linguistic feature. The second, technically more elaborate, half discusses factor analysis as a means for combining such counts across multiple features and arriving at more concise descriptors of differences across registers. Both questions are rephrased several times in different forms, involving lexical, syntactic, or discourse features and contrasts between speech and written text, between native and ESL speakers, and across different types of scientific writings, to name a few. In each case, the methodology behind corpus selection, feature identification, calculation of the results with a computer, manual editing, and interpretation of the results is explained in sufficient detail. There is a collection of highly readable, if somewhat short, essays on methodological issues at the end of the book; these range from corpus design to tagging to (very briefly) statistical interpretation. There is also a substantial bibliography on the subject (up to 1996).

The book appears to be aimed at undergraduates in linguistics with no experience of corpus linguistics; even moderately technical linguistic terms are explained upon first mention. It will also be useful to those theoretical linguists who have managed to remain unaware of quantitative methods during the past ten years. Each analysis is presented very clearly, and can be replicated by the reader (the authors provide an extensive list of easily accessible corpora and concordancing tools), enhancing the book's value

for the classroom. The authors make a particular effort to avoid intimidating the reader with computer terminology or program listings; not a single line of code appears in the book.

However, the book's scope is narrow. The reader familiar with Biber's earlier work and that of his students learns little new—the methodology is substantially the same as in Biber's earlier *Variation across Speech and Writing* (1988) and almost all experimental results have been published before. There is no discussion of predictive methods or of statistical models of language, both of which this reviewer would consider appropriate for a volume titled "Corpus Linguistics." Even in the area that the book does cover, several important issues receive inadequate treatment. There is only brief discussion and no detailed calculations of statistical tests of significance, so the (presumably inexperienced) reader has no means of deciding when the corpus and the feature counts are large enough. Even though factor analysis is the centerpiece of the second half of the book, there is no explanation of how the factors and their loadings might be obtained in practice. These omissions become more apparent because of the book's one shortcoming in presentation: although the authors are lucid, they tend to be repetitive as they detail only superficially dissimilar studies. The reader tends to search for nonexistent differences in methodology, and, after a few sections, is tempted to jump to the study's results and summary, which do make for interesting reading.

This book could be used as part of the curriculum for an introductory corpus linguistics class, especially for students with a humanities background. But it will probably provide only for a portion of a semester, and will not give by itself a sufficiently well-rounded perspective in corpus linguistics—don't throw out your copy of Mosteller and Wallace yet.—*Vasileios Hatzivassiloglou, Columbia University*

References

- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge University Press.
- Mosteller, Frederick and David L. Wallace. 1984. *Applied Bayesian and Classical Inference: The Case of The Federalist Papers*. (Second edition of *Inference and Disputed Authorship: The Federalist* [1964]). Springer-Verlag, New York.

Lexical Relations

Jean-Pierre Koenig

(State University of New York at Buffalo)

Stanford: CSLI Publications (Stanford monographs in linguistics), 1999, x+213pp; distributed by Cambridge University Press; hardbound ISBN 1-57586-177-1, \$49.95; paperbound, ISBN 1-57586-178-3, \$18.95

"Recent work in syntax has shown that much of a language's behavior stems from the information structure associated with its words. The thrust of this book is to provide a model of lexical relations which reconciles the lexicon's idiosyncratic and productive aspects. Building on work in Head-driven Phrase-Structure Grammar, an organization of lexical knowledge—called the Type Underspecified Hierarchical Lexicon—is proposed, through which partial regularities, medium-size generalization, and truly productive processes receive a unified model. Its basic thesis is that all lexical relations reduce to categorization (the membership of the two related lexemes in a common category) and that category intersection is the only mechanism needed to model lexical processes provided lexical items can be stored partially underspecified as to their category membership. Aside from the conceptual simplification that results from this move, the book demonstrates that several empirical and theoretical benefits accrue to this architecture; in particular, many salient properties of morphological processes are shown to reduce to inherent, formal properties of the organization of the lexicon."—*From the publisher's announcement*

Semantics and Syntax in Lexical Functional Grammar: The Resource Logic Approach

Mary Dalrymple (editor)

(Xerox Palo Alto Research Center)

Cambridge, MA: The MIT Press (Language, speech, and communication series), x+399 pp; hardbound, ISBN 0-262-04171-5, \$40.00

"A new, deductive approach to the syntax-semantics interface integrates two mature and successful lines of research: logical de-

duction for semantic composition and the Lexical Functional Grammar (LFG) approach to the analysis of linguistic structure. It is often referred to as the 'glue' approach because of the role of logic in gluing meanings together.

"The glue approach has attracted significant attention from, among others, logicians working in the relatively new and active field of linear logic; linguists interested in a novel deductive approach to the interface between syntax and semantics within a nontransformational, constraint-based syntactic framework; and computational linguists and computer scientists interested in an approach to semantic composition that is grounded in a conceptually simple but powerful computational framework.

"This introduction to and overview of the new approach is the first book to bring together the research of the major contributors to the field."—*From the publisher's announcement*

Evaluation of the Linguistic Performance of Machine Translation Systems: Proceedings of the [Workshop at] Konvens '98 in Bonn

Rita Nübel and Uta Seewald-Heeg (editors)

(Universität des Saarlandes and Universität Hannover)

St. Augustin, Germany: Gardez! Verlag (Sprachwissenschaft, Computerlinguistik und neue Medien, edited by Nico Weber, volume 2), 1998, 178 pp; paperbound, ISBN 3-928624-97-0, DM 49.90

The contents of the volume are as follows:

"Linguistically oriented evaluation of machine translation" by Rita Nübel and Uta Seewald-Heeg

"Textsortenspezifische Evaluation maschineller Übersetzungssysteme am Beispiel von Instruktionstexten" by Uta Seewald-Heeg

"Phänomenspezifische Evaluierung von maschinellen Übersetzungssystemen am Beispiel von Koordination" by Rita Nübel

"Komposita im Internet—Übersetzungssysteme auf dem Woodway" by Ulrike Ulrich

"Zur Problematik der maschinellen Übersetzung von Nebensätzen zwischen den

Sprachen Englisch und Deutsch" by
Stephan Mehl, Britta Heidemann, and
Martin Volk
"Small scale evaluation methods" by
Simone Wagner
"DiET in the context of MT evaluation" by
Judith Klein, Sabine Lehmann, Klaus
Netter, and Tillmann Wegst
"Bewertung von MT-Systemen aus

Benutzersicht: Evaluierung im Projekt
MIROSLAV" by Jutta Marx
"Evaluation problems from a developer's
point of view" by Kirsten Falkedal
"Evaluation design: The EAGLES frame-
work" by Margaret King
"Linguistic MT evaluation: Assessment and
summary" by Rita Nübel and Uta
Seewald-Heeg