

Briefly Noted

Une grammaire électronique du français

Anne Abeillé

(Université Denis Diderot (Paris VII))

Paris: CNRS Éditions (Collection sciences du langage, edited by Christian Hudelot), 2002, 396 pp; paperbound, ISBN 2-271-05824-4, €38.00

This recent publication by CNRS Publications is a reworked version of Abeillé's 1991 doctoral dissertation. It presents, in considerable detail, the tree adjunction grammar (TAG) formalism used, and the details of the French grammar written in that formalism. This book is worthy of notice because of its thoroughness in presenting a complete grammar of French for computational use, though, of course, the content is accessible only to readers of French.

Chapter 1 is an overview of the TAG framework used throughout the book. Chapter 2 discusses the basic properties of French and justifies the fact that the analyses do use a VP. Chapter 3 is an overview of elementary trees and lexical categories. Chapter 4 discusses further categories of elementary trees. Chapter 5 addresses long-distance dependencies and extraction phenomena. Chapter 6 discusses noncompositional expressions. Chapter 7 covers implementation issues and some testing results. The book ends with a conclusion and two appendices, one on the formal properties of TAG and one on French TAG grammar.

The overview of the TAG formalism is very well done. It is pedagogically sound, and French speakers would find it a good overview of pertinent work and of the aspects of a TAG relevant for a grammar writer. Although a novice in the TAG framework, I found myself well prepared to understand the specifics of the following chapters devoted to French constructions.

The sections on French-specific issues (auxiliaries, clitics, past-participle agreement) and the presentation of the elementary trees are both clear and illustrate well how the framework can be used to generate the right results. The details are complex and will require an advanced understanding of the linguistic issues of French syntax. One naturally makes allowances for the fact that the work covers all of the syntax of French, no small task.

The impression that emerges from reading this book is that creating a grammar in TAG would require a lot of effort—for the trees as well as for the complex lexical items. This seems to be borne out by the relatively small lexicon (7,500 items) that the system uses (p. 289), even though the development time spans over 10 years. Size, performance, and coverage, though briefly documented, also seem to be of lesser importance, as this is a theoretical implementation as much as, or more than, a usable computational grammar.

A detail that I find unconvincing from a linguistic perspective is the account of island constraints. The island nature of a complement clause is accounted for (p. 239) by stating that it results from a substitution operation rather than attachment to a foot node (*nœud pied*). Since no reference to previous work appears here, one is led to believe this is a new explanation, and one that struck me as mechanical and unconvincing.

From a practical point of view, I cannot agree with the insistence, led by theoretical concerns, that the grammar be generative (i.e., able to reject ungrammatical input). My work on grammar in industrial contexts has convinced me that “real” corpora are flawed in form and rarely conform to the theoretical assumptions of grammar writers. I would have welcomed some estimation of how this grammar might apply in real contexts.—*Jessie Pinkham, Microsoft Research*

Natural Language Processing for Online Applications: Text Retrieval, Extraction, and Categorization

Peter Jackson and Isabelle Moulinier

(Thomson Legal & Regulatory)

Amsterdam: John Benjamins Publishing Company (Natural language processing series, edited by Ruslan Mitkov, volume 5), 2002, x+225 pp; hardbound, ISBN 90-272-4988-1 and 1-58811-249-7, \$68.00, €75.00; paperbound, ISBN 90-272-4989-X and 1-58811-250-0, \$29.95, €33.00

Jackson and Moulinier's book introduces applied natural language processing to an audience that need not have any prior knowledge of the field or of linguistic issues. But it is not *NLP for Dummies*; it respects the intelligence of its readers (apart from a few too-glib jokes in the first chapter) and doesn't

shy away from serious mathematics where the treatment warrants it.

The introductory chapter covers the goals of NLP and fundamental tools such as tokenizers, part-of-speech taggers, and simple name recognizers. The four main chapters of the book then cover document retrieval from collections and from the Web, information extraction, text categorization, named-entity and coreference recognition, and automatic summarization.

Thus, unlike the standard large and comprehensive computational linguistics textbooks (Jurafsky and Martin 2000; Manning and Schütze 1999) and handbooks (Dale, Moisl, and Somers 2000; Mitkov 2003), Jackson and Moulinier's book is concise and focused on the goal of building applications. The book is therefore a useful resource for those who want to find out quickly about NLP without learning everything there is to know about computational linguistics and NLP. A reader who completes the book will be well-equipped then to learn more by reading selectively from the textbooks and handbooks.

Readers of this journal who work in industry will want to recommend the book to coworkers, such as project managers and software engineers, whose experience is in other fields; those who work in universities will want to recommend it as prereading to keen undergraduates. And any reader of this journal will find it a helpful consolidation of up-to-date material that is presently scattered around conference proceedings and journal articles.—*Graeme Hirst, University of Toronto*

References

- Dale, Robert, Hermann Moisl, and Harold Somers, editors. 2000. *Handbook of Natural Language Processing*. Marcel Dekker, New York.
- Jurafsky, Daniel and James Martin. 2000. *Speech and Language Processing*. Prentice-Hall, Englewood Cliffs, NJ.
- Manning, Christopher and Hinrich Schütze. 1999. *Foundations of Statistical Natural*

Language Processing. MIT Press, Cambridge.

Mitkov, Ruslan, editor. 2003. *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford.

Defining Language: A Local Grammar of Definition Sentences

Geoff Barnbrook
(University of Birmingham)

Amsterdam: John Benjamins (Studies in corpus linguistics, edited by Elena Tognini-Bonelli, volume 11), 2002, xv+280 pp; hardbound, ISBN 90-272-2281-9 and 1-58811-298-5, \$79.00, €88.00.

"This book describes the analysis of the main features of the language used in English definition sentences, using as a corpus the definitions contained in the *Collins Cobuild Student's Dictionary*. It examines the usefulness of the information provided by dictionaries in natural language processing work and the nature of the language used in dictionary definitions in general and in the Cobuild range in particular. It provides a general survey of monolingual English dictionaries, including a brief history of their development, and a detailed investigation of the nature of learners' dictionaries and their special features. The concept of sublanguages is examined, together with the justification for regarding definition sentences as a sublanguage and for the application to them of a local grammar of definition. Grammars and parsers are considered in general terms, and in their relevance to the creation of a model for the language of definitions.

"The methodology adopted for the development of the language model is described, together with a detailed account of the taxonomy, local grammar, and associated parser developed for definition sentences. The implications of the results of the analysis and future possible applications of the taxonomy, grammar, and parser are described and assessed."—*From the précis of the book*