

# Briefly Noted

## Les linguistiques de corpus

**Benoît Habert, Adeline Nazarenko, and André Salem**

(ENS de Fontenay-Saint-Cloud, Université Paris-XIII, and Université Paris-III)

Paris: Editions Armand Colin, 1997, 240 pp; paperbound, ISBN 2-200-01775-8, FF 125.00

In *Les linguistiques de corpus*, newcomers as well as experts in the field will find a very interesting, well-written, and concise survey of corpus linguistics. The book gives a good overview of the multiple methods used and the problems involved in the creation and the annotation of a corpus. It also presents proposed applications for these corpora. The authors do not relate their own work, but present to the reader in an organized way (by topic) the work of many researchers. The book is written in French, but most of the works presented are based on corpora of English texts.

Among other things, the authors discuss what makes a "good corpus," how representative and reliable a corpus can be, and what corpora annotated at the lexical, syntactic, or semantic level presently exist. They analyze the interesting problem of the dependency of a corpus annotation on the foreseen applications for the corpus. The authors present applications such as word sense disambiguation and lexical acquisition from text, but they also present different applications that are of more concern to the literary community, such as discovering changes in the usage of a word between periods or between authors.

In the authors' view (which I certainly share and consider very positive), corpus linguistics leaves linguists (and language specialists) and computer scientists no choice but to finally come together and use each other's strengths (deep understanding of language and deep understanding of computer technology); they must share their expertise to push the field forward.—*Caroline Barrière, School of Information Technology and Engineering, University of Ottawa, Canada*

## Text Retrieval and Filtering: Analytic Models of Performance

**Robert M. Losee**

(University of North Carolina)

Boston: Kluwer Academic Publishers (The Kluwer international series on information retrieval, edited by W. Bruce Croft), 1998, x+242 pp; hardbound, ISBN 0-7923-8177-7, \$115.00, £78.25, Dfl 260.00

This is the first book that addresses the problem of analytically computing the performance of text retrieval and filtering systems. It describes means by which retrieval may be studied analytically, allowing one to describe current performance, to predict future performance, and to understand why systems perform as they do. The focus is on retrieving and filtering natural language text—full sentences as well as phrases and individual words. The book addresses retrieval performance for the simple case of queries with a single term, for the more complex case with multiple terms (both with term independence and term dependence), and for the use of grammatical information to improve performance. Unambiguous statements of the conditions under which one method or system will be more effective than another are developed. The last chapter explicitly addresses how grammatical constructs and methods may be studied in the context of retrieval or filtering system performance. The book builds towards solving this problem, although the material in earlier chapters is as useful to those addressing nonlinguistic, statistical concerns as it is to linguists.—*Based on the publisher's announcement*

## Linguistic Specifications for Typed Feature Structure Formalisms

Frank Van Eynde and Paul Schmidt  
(editors)

(Katholieke Universiteit Leuven and Universität des Saarlandes)

Luxembourg: Office for Official Publications of the European Communities (Studies in machine translation and natural language processing, edited by Erwin Valentini, volume 10), 1998, 344 pp; paperbound, ISSN 1017-6568, ECU 15.00

"This volume is an abridged and revised version of the final report of the project "Investigation of Linguistic Specifications for Future Industrial Standards". This project was part of the Multilingual Action Plan of the European Union (1994–1995) and was carried out by a consortium of seven research institutes. . . .

"The purpose of the project was the development of linguistic specifications for TFS-based formalisms, i.e. formalisms which make use of typed feature structures (TFS) for the representation of linguistic information. . . . These specifications are not geared to any particular language, but provided a starting point for the development of more detailed specifications for a number of individual languages. . . . The linguistic specifications which will be presented in this vol-

ume . . . aim at compatibility with any kind of TFS-based formalism.

"Apart from compatibility with this family of NLP formalisms there are a number of further requirements which the specifications aim to fulfill, such as breadth of coverage, internal coherence and multilingual orientation. On a more technical level the specifications are required to be monostratal, lexicalist and constraint-based. Since these are the main characteristics of Head-driven Phrase Structure Grammar, it will not come as a surprise that HPSG forms the point of departure for most of the work which will be presented in this volume."—*From the editors' introduction*

The contents of the volume are as follows:

Introduction by Frank Van Eynde and Paul Schmidt;

"Formal assumptions" by Paul Schmidt;

"Lexical generalisations" by Stella Markantonatou and Louisa Sadler;

"Phrase structure" by Paul Bennett and Paul Schmidt;

"Predicate-argument structure" by Toni Badia and Carme Colominas;

"Tense, aspect and negation" by Frank Van Eynde;

"Determination and quantification" by Valerio Allegranza;

"Support verb constructions" by Fiammetta Namer.