Briefly Noted

Natural Language Processing Using Very Large Corpora

Susan Armstrong, Kenneth Church, Pierre Isabelle, Sandra Manzi, Evelyne Tzoukermann, and David Yarowsky (editors)
(University of Geneva, AT&T Labs–Research, Xerox Research Centre Europe, University of Geneva, Bell Laboratories, and Johns Hopkins University)

This volume is a collection of important papers selected from the proceedings of the first three Workshops on Very Large Corpora (Columbus, 1993; Kyoto, 1994; Cambridge, MA, 1995) and the 1995 workshop entitled From Text to Tags (Dublin). The editors write in their introduction: "The success of these workshops was in some measure a reflection of the growing popularity of corpus-based methods in the NLP community. But first and foremost, it was due to the fact that the workshops attracted so many high-quality papers. The importance of this material for the field is such that it deserves to be made more readily available than harder-to-find or out-of-print workshop proceedings." The contents of the volume are as follows:

"Implementation and evaluation of a German HMM for POS disambiguation" by Helmut Feldweg
"Improvements in part-of-speech tagging with an application to German" by Helmut Schmid
"Unsupervised learning of disambiguation rules for part-of-speech tagging" by Eric Brill and Mihai Pop
"Tagging French without lexical probabilities—Combining linguistic knowledge and statistical learning" by Evelyne Tzoukermann, Dragomir Radev, and William Gale
"Example-based sense tagging of running Chinese text" by Xiang Tong, Chang-ning Huang, and Cheng-ming Guo
"Disambiguating noun groupings with respect to WordNet senses" by Philip Resnik
"A comparison of corpus-based techniques for restoring accents in Spanish and French text" by David Yarowsky
"Beyond word *N*-grams" by Fernando Pereira, Yoram Singer, and Naftali Tishby
"Statistical augmentation of a Chinese machine-readable dictionary" by Pascale Fung and Dekai Wu
"Text chunking using transformation-based learning" by Lance Ramshaw and Mitchell P. Marcus
"Prepositional phrase attachment through a backed-off model" by Michael Collins and James Brooks
"On the unsupervised induction of phrase-structure grammars" by Carl de Marcken
"Robust bilingual word alignment for machine aided translation" by Ido Dagan, Kenneth Church, and William Gale
"Iterative alignment of syntactic structures for a bilingual corpus" by Ralph Grishman
"Trainable coarse bilingual grammars for parallel text bracketing" by Dekai Wu
"Comparative discourse analysis of parallel texts" by Pim van der Eijk
"Comparing the retrieval performance of English and Japanese text databases" by Hideo Fujii and W. B. Croft
"Inverse document frequency (IDF): A measure of deviations from Poisson" by Kenneth Church and William Gale

Techniques in Speech Acoustics

Jonathan Harrington and Steve Cassidy
(Macquarie University)

"This book is the development of a series of lectures to undergraduate and post-graduate students at Macquarie University on basic principles in acoustic phonetics and speech signal processing. The first part of the book is intended to provide students with the ability to interpret acoustic records of speech signals in their various forms. These chapters include a review of elementary wave motion and frequency analysis as applied to speech, a summary of the relationship between speech production and its acoustic consequences, and a review of the principal cues to speech sounds and prosodic units.

The material from these first four chapters (and the related exercises on the accompanying CD-ROM) has formed the basis of a one-semester undergraduate course in acoustic phonetics to students primarily of linguistics, but also of other disciplines including computer science and psychology.

"The second part of the book provides an introduction to speech signal processing, which is intended for similar groups of students. It is therefore different from more detailed introductory texts in this area, which assume both a background in engineering / signal processing and a more sophisticated mathematical knowledge. Part of the motivation for writing this section of the book is to make many of the techniques and algorithms that are discussed in the engineering literature on speech analysis more accessible to both students and researchers of phonetics and speech science whose training is not usually in a scientific discipline. We have therefore tried to keep equations to a minimum and to assume, as far as possible, no more than a very basic understanding of algebra and trigonometry. In this part of the book we cover fundamental aspects of time and frequency domain processing of speech signals, digital techniques for combining (in digital formant synthesis) and separating (in linear predictive coding) the contributions of the source and filter to the acoustic speech signal, and techniques for the probabilistic classification of acoustic speech data that form the basis of more advanced work in automatic speech recognition."—*From the preface (abridged)*

Syntactic Theory: A Formal Introduction

Ivan A. Sag and Thomas Wasow
(Stanford University)

Stanford, CA: CSLI Publications (CSLI lecture notes, number 92), 1999, xiii+481 pp; distributed by Cambridge University Press; hardbound, ISBN 1-57586-161-5, $74.95; paperbound, ISBN 1-57586-160-7, $29.95

"*Syntactic Theory: A Formal Introduction* ... marks a return to 'generative grammar' in its original sense. This book focuses on the development of precisely formulated grammars whose empirical predictions can be directly tested. There is considerable emphasis on prediction and evaluation of grammatical hypotheses, as well as on integrating syntactic hypotheses with matters of semantic analysis. Problem solving is also emphasized; the extensive problem sets draw from a variety of languages other than English. Special attention is paid to the nature of lexical entries and the organization of the lexicon in terms of type hierarchies and constraint inheritance. The theoretical perspective of the book is presented in the context of current models of language processing, which provide motivation for a constraint-based, lexicalist grammatical architecture, whose value has already been demonstrated in computer language processing applications."—*From the publisher's announcement*

Word-Order Based Grammar

Eva Koktova

Berlin: Mouton de Gruyter (Trends in linguistics: studies and monographs, edited by Werner Winter, volume 121), 1999, xv+389 pp; hardbound, ISBN 3-11-061252-0, DM 198.00

"In this book, I propose, in outline, a new theory of grammar, which I call Word-Order Based Grammar.

"In my theory, word order is the primary grammatical principle. It is accompanied by the principle of lexical valency, and the principle of the recursive expansion of the sentence.

"I offer alternative accounts of the major problems of linguistic theory in terms of word-order: for the scoping properties of the operators of natural language ...; for anaphoric reference ...; for *wh*-extraction, movement and parasitic gaps ...; and for word-order in general ...

"I propose that there are five basic types of word-order: fixed deep word-order, free deep word-order, fixed surface word-order, free surface word-order corresponding to deep word-order, and free surface word-order not corresponding to deep word-order."—*From the introduction*

Computational Linguistics in the Netherlands 1998: Selected Papers from the Ninth CLIN Meeting

Frank Van Eynde, Ineke Schuurman, and Ness Schelkens (editors)

The volume contains revised versions of 14 papers that were first presented at the ninth annual Computational Linguistics in the Netherlands conference, held in Leuven, Belgium, in December 1998. The contents of the volume are as follows:

"Instance families in memory-based language learning" by Antal van den Bosch

"On the arbitrariness of lexical categories" by Gert Durieux, Walter Daelemans, and Steven Gillis

"An FGREP investigation into phonotactics" by Peter Kleiweg and John Nerbonne

"Exploring phonotactics with simple recurrent networks" by Ivelin Stoianov and John Nerbonne

"Style adaptation of statistical language models" by Dong Hoon Van Uytsel, Patrick Wambacq, and Dirk Van Compernolle

"Memory-based word sense disambiguation" by Jorn Veenstra, Antal van den Bosch, Sabine Buchholz, Walter Daelemans, and Jakub Zavrel

"The Amazon grammar and the last part of the middle field" by Simon van Dreumel

"Position of Polish clitics: An HPSG approach" by Anna Kupść

"Null-headed nominals in German and English" by John Nerbonne and Tony Mullen

"Corpus-based analysis of parenthetical reporting clauses" by Carla Schelfhout

"Figure heads in HPSG" by Frank Van Eynde

"A modern computational linguistics course using Dutch" by Gosse Bouma

"XML mixed content grammars" by Pim van der Eijk and Dennis Janssen

"Evaluation of the NLP components of the OVIS2 spoken dialogue system" by Gert Veldhuijzen van Zanten, Gosse Bouma, Khalil Sima'an, Gertjan van Noord, and Remko Bonnema