Briefly Noted

Computers, Language and Speech: Formal Theories and Statistical Data

Karen I. B. Spärck Jones, Gerald J. M. Gazdar, and Roger M. Needham (editors)
(University of Cambridge, University of Sussex, and Microsoft Research Ltd)

Thirteen invited papers from a two-day discussion session held in September 1999 are collected in this issue of *Philosophical Transactions of The Royal Society*. All the papers have been revised and edited, and the discussion following the presentation of each paper has been transcribed and appended. In addition, the editors have contributed a paper-length introduction. The contents are as follows:

"Introduction: Combining formal theories and statistical data in natural language processing" by Karen I. B. Spärck Jones, Gerald J. M. Gazdar and Roger M. Needham

"Formal grammar and information theory: Together again?" by Fernando Pereira

"Finite-state models, event logics and statistics in speech recognition" by Julie Carson-Berndsen

"Statistical and logical reasoning in disambiguation" by Stephen G. Pulman

"Towards a psycholinguistic computational model for morphological parsing" by R. Harald Baayen and Robert Schreuder

"Information extraction from broadcast news" by Yoshihiko Gotoh and Steve Renals

"Incorporating linguistic structure into statistical language models" by Ronald Rosenfeld

"Incorporating linguistic theories of pronunciation variation into speech-recognition models" by Mari Ostendorf

"The role of taxonomy in language engineering" by Geoffrey Sampson

"Learning dependency transduction models from unannotated examples" by Hiyan Alshawi and Shona Douglas

"Stochastic text generation" by Jon Oberlander and Chris Brew

"Probabilistic methods in spoken-dialogue systems" by Steve J. Young

"Concept-to-speech synthesis by phonological structure matching" by P. A. Taylor

"Prosody modelling in concept-to-speech generation: Methodological issues" by Kathleen R. McKeown and Shimei Pan

The CHILDES Project: Tools for Analyzing Talk (third edition)
Volume I: Transcription format and programs
Volume II: The database

Brian MacWhinney
(Carnegie Mellon University)

"[These] two volumes . . . document the three components of the CHILDES Project. [The first] is divided into two parts that provide an introduction to the use of computational tools for studying language learning. The first part is the CHAT manual, which describes the conventions and principles of CHAT transcription and recommends specific methods for data collection and digitization. The second part is the CLAN manual, which describes the uses of the editor, sonic CHAT, and the various analytic commands. The book will be useful for both novice and experienced users of the CHILDES tools, as well as instructions and students working with transcripts of child language.

"The second volume of this set describes in detail all of the corpora included in the CHILDES database. The conversational interactions in the corpora come from monolingual children and their caregivers and siblings, as well as bilingual children, older school-aged children, adult second-language learners, children with various types of language disabilities, and aphasics recovering from language loss. The database includes transcripts in 26 different languages.

"The CD-ROM that accompanies these volumes includes the transcript files described in Volume II. It runs on both Windows and Macintosh platforms."—*From the publisher's announcement*

Unicode^{TM}: A Primer

Tony Graham
(Mulberry Technologies Inc)

"Unicode is a character encoding standard that includes all of the major scripts of the world in a simple and consistent manner. It is an integral part of HTML 4.0 and XML .... This book is a primer for Unicode and its uses.... [It] doesn't attempt to replicate the Unicode Standard. Instead, it ... introduces the principles of Unicode, describes both the Unicode Standard and the other material from the Unicode Consortium not included in the standard, [and] shows how the Unicode Standard is used on the Internet, in operating systems, and in programming languages."—*From the author's preface*