# Book Reviews

## Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition

**Daniel Jurafsky and James H. Martin**
(University of Colorado, Boulder)

Upper Saddle River, NJ: Prentice Hall
(Prentice Hall series in artificial
intelligence, edited by Stuart Russell
and Peter Norvig), 2000, xxvi+934 pp;
hardbound, ISBN 0-13-095069-6, $64.00

*Reviewed by*
*Virginia Teller*
*Hunter College and the Graduate Center,*
*The City University of New York*

### 1. Introduction

Jurafsky and Martin's long-awaited text sets a new gold standard that will be difficult to surpass, as attested by the flurry of glowing reviews that accompanied its publication early this year.[1] In a nutshell, J&M have successfully integrated knowledge-based and statistical methods with linguistic foundations and computational models in a book that is at once balanced, comprehensive, and, above all, eminently readable. My review will focus primarily on the format and content of *Speech and Language Processing* (*SLP*) as well as its use in a course.

### 2. Format

The 21 chapters are grouped into an introduction followed by four parts starting at the level of Words (six chapters), then proceeding to Syntax (six chapters), Semantics (four chapters), and finally Pragmatics (four chapters). Several other writers contributed chapters, principally Andrew Kehler (Chapter 18, "Discourse"), Keith Vander Linden (Chapter 20, "Natural Language Generation") and Nigel Ward (Chapter 21, "Machine Translation"). Supplementary material in four appendices covers regular-expression operators, the Porter stemming algorithm, the CLAWS C5 and C7 tagsets, and the forward-backward algorithm.

Instructional aids abound. Margin notes are used liberally, and a series of Methodology Boxes explains techniques for evaluating natural language systems, including such measures as perplexity for *n*-gram models, the kappa statistic for computing agreement, and precision and recall for parsing and information extraction and

1 See, for example, the book's page at amazon.com or the book's home page at www.cs.colorado.edu/~martin/slp.html.

retrieval. So-called Key Concepts highlight important points, and each chapter ends with a summary, bibliographic and historical notes, and exercises. In addition there is an extensive bibliography (more than 50 pages) and index (over 30 pages).

Pithy epigrams introducing each chapter and sprinkled throughout the text further enliven J&M's engaging writing style. Two of my favorites appear in Chapter 17 ("Word Sense Disambiguation and Information Retrieval"), which opens with Groucho Marx (*Oh, are you from Wales? Do you know a fella named Jonah? He used to live in whales for a while.*), and Chapter 19 ("Dialogue and Conversational Agents"), which features Abbott and Costello's version of *Who's on First*. Excerpts are also taken from poems (Frost), musicals (Gershwin, Gilbert and Sullivan), drama (Shakespeare), and literature (Bacon, Melville), and many leading figures in the field—Chomsky, Jespersen, and Jelinek, to name a few—are quoted as well.

**3. Content**

*SLP* delves into topics as diverse as articulatory phonetics and the Chomsky hierarchy and treats virtually every major application in the field. I can't think of an area that's not covered. Spelling and grammar correction, speech recognition, text-to-speech, tagging, context-free and probabilistic parsing, word sense disambiguation, dialogue and conversational agents, information extraction and retrieval, natural language generation, machine translation, and many more are all included. On the theoretical side, J&M discuss regular expressions, finite-state automata and transducers, first-order predicate calculus, grammar formalisms, and more.

The material on each subject is presented in a logical and organized fashion. Theoretical underpinnings are explored in depth, computational modeling strategies are thoroughly motivated, and algorithms are clearly stated and illustrated with detailed examples. Even areas that are touched upon only lightly, such as word segmentation (Chapter 5) and computational approaches to metaphor and metonymy (Chapter 16), are mentioned with sufficient pointers to the literature that they can easily be pursued by interested readers.

A particular strength of *SLP* is the way that central themes are woven into the fabric of specific applications. The noisy-channel model is introduced in Chapter 5 ("Probabilistic Models of Pronunciation and Spelling"), and the simplified version of Bayes's rule as the product of likelihood and prior probability is stated as a Key Concept (p. 149). J&M apply these notions to spelling correction and English pronunciation variation in Chapter 5. The noisy-channel model is then reformulated for speech recognition in Chapter 7, and the revised version of Bayes's rule appears again as a Key Concept (p. 239). Bayesian inference resurfaces briefly in Chapter 17 as the naive Bayes classifier supervised learning method for word sense disambiguation, and Bayes's rule is restated a final time in the context of statistical machine translation (Chapter 21).

The dynamic programming paradigm traverses a similar course, beginning in Chapter 5 with the minimum edit-distance, forward, and Viterbi algorithms. The Viterbi algorithm is revisited in Chapter 7 and used, along with A* decoding, for speech recognition. Chapters 10 and 12 describe the Earley and CYK algorithms for context-free and probabilistic context-free parsing, respectively, and Appendix D sketches the forward-backward (Baum-Welch) algorithm, a special case of the EM algorithm. This threading of themes throughout the text, rather than seeming disjointed, permits J&M to consider applications in their proper context (words, syntax, semantics, pragmatics), and they provide ample explicit links along the way.

As is typical with the first printing of any new book, there are numerous typographical and other printing errors—some more serious than others—that readers

may find annoying. J&M have compiled a list of errata, currently about five pages, and posted it on the *SLP* web site. At this writing (July 2000), a second printing incorporating all of these corrections is expected shortly.

## 4. Use in a Course

In early 2000, I used *SLP* to teach an introductory graduate course to linguistics and computer science students at the CUNY Graduate Center; linguists outnumbered computer scientists two to one. The computational sophistication of the linguists varied from none to intermediate C++ programming skills, while the other students had minimal, if any, background in linguistics.

Copies of the book arrived in the college bookstore just one day before the first class. During the course of the 14-week semester we covered 12 chapters, approximately one chapter per week and roughly three chapters from each of the book's four parts; the last two weeks were devoted to student presentations (see below). Homework assignments took two forms: chapter exercises and Web-based work. Whenever feasible, chapter exercises involving computation, such as computing minimum edit-distance or using kappa to determine the agreement between the output of two taggers, were selected so that students with programming skills could obtain solutions by writing programs, while students without this ability could perform the same calculations by hand. The Web assignments allowed students to construct their own corpora and then evaluate the results of processing them using taggers, parsers, and machine translation. Other Web sites gave exposure to natural language generation and to lexical semantics via WordNet. Unlike other texts I have used for this course (see Teller 1995), no supplementary reading is required with *SLP*. For content, it's one-stop shopping.

Perhaps the best indication of *SLP*'s worth came from the research component of the course. Students picked their own topics and developed a project in five steps: annotated bibliography, abstracts of relevant literature, proposal, class presentation, and final written report. This approach gave the students an opportunity to tackle a variety of problems close to their own interests. Almost everyone chose to use corpus-based methods, which entailed building a corpus from Web sources and using it as primary data. Here's a sample of topics: the productivity of two similar Chinese affixes (*zi* and *tou*); English adjective usage for Japanese learners; word sense disambiguation in medical text; natural language information retrieval on the Web; unknown words in speech recognition; modifying Brill's tagger for e-mail. The students were uniformly enthusiastic about *SLP* as a text for the course, and they also credited it as a valuable starting point for their research. I credit *SLP* for contributing to the great distance in knowledge that some members of the class were able to travel from the beginning to the end of the semester.

In their Preface, J&M suggest four routes through *SLP* for courses with different foci: NLP (one quarter); NLP (one semester); Speech plus NLP (one semester); and computational linguistics (one quarter). Compared to Allen (1995), *SLP* conveys a more intuitive grasp of human language, offers superior coverage of statistical methods, and provides a more up-to-date treatment of the symbolic paradigm. Although *SLP* may not suffice for a course that concentrates on the stochastic approach alone, some instructors may nonetheless find J&M's tutorial level of exposition more suitable than either Charniak's (1993) terse account or Manning and Schütze's (1999) monumental tome. I plan to use *SLP* for an advanced undergraduate computer science course at Hunter College in late 2000.

## 5. Conclusion

In the past decade there has been a sea change in the methodology of natural language computing and an explosion in the availability of applications, especially on the Web. Traditional labels such as *computational linguistics* and *natural language processing* no longer adequately describe the range of methodologies and applications in the field; today a better term is *language technology*. *SLP* is the first text to address the needs of the wide audience in this expanded arena. Readers familiar with linguistics will find an accessible introduction to computational modeling, methods, algorithms, and techniques. Those with a computer science background will gain insight into the linguistic foundations of the field. And people knowledgeable about speech processing can learn more about syntax, semantics, and discourse. The appeal of *SLP* is by no means limited to academia. Researchers, developers, and practitioners from any related discipline should all find *SLP* an ideal vehicle for becoming acquainted with the burgeoning field of language technology.

**References**

Allen, James. 1995. *Natural Language Understanding*, second edition. Benjamin/Cummings, Redwood City, CA.

Charniak, Eugene. 1993. *Statistical Language Learning*. The MIT Press, Cambridge, MA.

Manning, Christopher D. and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, MA.

Teller, Virginia. 1995. Review of *Natural Language Processing* (second edition) by James Allen. *Computational Linguistics* 21(4):604–606.

*Virginia Teller* is Professor of Computer Science at Hunter College, City University of New York, and a member of the doctoral faculties in Linguistics and Computer Science at the CUNY Graduate Center. In recent years her research has focused on machine translation and parameter-based models of first-language syntax acquisition. Teller's address is: Computer Science Department, Hunter College, CUNY, 695 Park Avenue, New York, NY 10021; e-mail: teller@cs.hunter.cuny.edu.