chart parsing is reduced into a set of matrix operations dealing with sparse matrices. Appendices list sample grammars and lexicons, which brings substance to the claims.

"Speech" in this book refers to English only, which is never made explicit. This seems to be the normal case in American literature, however. Of course, most of the contribution is relevant to other languages as well.

The book also provides an interesting contribution in the area of finite-state properties of language, because the phrase structure grammars used are essentially finite-state. Other finite-state accounts (such as the two-level model by Koskenniemi [1984] and cascaded transducers by Kaplan and Kay [Kay 1983] seem to have been less successful in combining structural information with segmental processes. Both other models are purely segmental, although syllables are sometimes referred to as contexts.

An interesting problem concerning rule interaction in the proposed formalism is dealt with on page 113. There would be an obvious need for subtraction (for defining negative contexts) and intersection (combining effects). Subtraction, however, turns out to exclude too much, whereas intersection is too permissive.

The book is well written and the argumentation proceeds logically. Both strong and weak points of the theories proposed are clearly presented. It gives a fair overall picture of the field of speech recognition, and much of the book could be suitable as a textbook. Nevertheless, some passages address mostly readers with a considerable background. The main topic covers, of course, a specific slice of the whole field, namely the treatment of allophonic variation. One minor inconvenience is the use of a reference format that cites a number only, not the author and the year. This results in a small savings in space but a larger burden for the reader. The book appears to be Church's (previously unpublished) doctoral dissertation from MIT, though this is not clearly indicated in the volume. Although not particularly new, it still is very valuable.

References

Kay, M. 1983 "When Meta-Rules Are not Meta-Rules," In Sparck Jones, K. and Wilks, Y. (eds.), *Automatic Natural Language Parsing.* Ellis Horwood, Chichester, U.K.

Koskenniemi, K. 1984 "A General Computational Model for Word-Form Recognition and Production," In *COLING '84: Proceedings of the 10th International Conference on Computational Linguistics,* Stanford, 178–181.

Kimmo Koskenniemi's Ph.D. thesis introduced a finite-state two-level model to account for inflection, derivation, and compounding in an efficient and language-independent way. He works on finite-state morphology and syntax at the Research Unit for Computational Linguistics at the University of Helsinki. Koskenniemi's address is: Department of General Linguistics, University of Helsinki, Hallituskatu 11, SF 00100 Helsinki, Finland. E-mail: koskenniemi@opmvax.csc.fi

---

Prosody and Speech Recognition

Alex Waibel
(Carnegie Mellon University)

San Mateo, CA: Morgan Kaufmann Publishers and London: Pitman, 1988, xii + 212 pp.
(Research Notes in Artificial Intelligence)
Softbound, ISBN 0-934613-70-2 and 0-273-08787-8, $22.95

Reviewed by
Joan Bachenko
AT&T Bell Laboratories

Most research in natural language processing (NLP) concentrates on the syntax and semantics of written language, a situation that exists in part because most NLP applications are concerned with systems that rely on written language analysis, e.g. information retrieval and text-generation systems. Recently, however, we have begun to see a growing interest in spoken language and the application of natural language processing to text-to-speech synthesis and speech recognition. Waibel's volume, which describes new results in automated speech recognition, makes an important contribution to this research direction.

At present, speech recognition technology gives us two choices: speaker-independent systems that handle small vocabularies (one to five phonetically distinct words) and require no training, or speaker-dependent systems that recognize somewhat larger vocabularies (up to 1,000 words online) and require training sessions for each user. Although experimental systems can recognize limited continuous speech, freely generated phrases and sentences cannot be processed, nor can words that are "unknown" to a recognizer. Waibel believes that this condition can change if recognition systems, which currently focus on identifying acoustic phonetic segments, are expanded to include prosodic information, i.e. information about nonsegmental features such as duration and pitch. His central claim is that a system equipped with prosody rules can achieve very large vocabulary recognition, up to 20,000 words, in both continuous speech and isolated word tasks. To make his point, Waibel examines four prosodic features—duration, intensity, pitch, and stress. For each feature, he discusses in detail a series of experiments that demonstrate the techniques that he used (and, in some cases, invented) for extracting the prosodic features, and rules that use prosodic feature patterns to narrow the search space for word hypothesization to a small subset of the total vocabulary. Waibel's results make a convincing case that prosody can play a valuable role in machine perception of speech; however, they fall short of establishing his strongest claims, as he lacks a complete implementation of the system.

Most of the book is organized around each of the prosodic features that Waibel investigates. His explication is

generally very thorough, although the chapter on pitch is unusually skimpy. For some reason, Waibel chooses to limit his discussion to two sentence tunes—yes-or-no questions (with a rising pitch contour) and declarative statements (falling pitch contour)—that he feels are relevant to connected speech recognition. Waibel does not investigate this in any depth, however; nor does he consider the possible contribution of pitch to the recognition of polysyllabic words, which themselves can make up a single intonation phrase.

The discussion of other prosodic features is much more satisfying, and I was especially intrigued by the chapter on stress. Previous researchers on speech recognition and prosody (e.g. Lea 1980) have held that stressed syllables provide "islands of phonetic reliability" because stress in itself makes a speech segment more identifiable to human and machine listeners. Waibel's results contradict this claim. His experiments show no significant effect of stress on recognition accuracies. What makes stressed syllables special is their closeness to lexical representation; stressed syllables have closer agreement between acoustic reality and abstract representation than unstressed syllables. A recognizer will therefore have better success at using stress to find syllables and words than particular phonetic segments; in Waibel's implementation, stress (specifically, stress probabilities) is used for locating word boundaries and for distinguishing between function and content words. Such results have important implications for the psychological reality of lexical representations and for the place of cognitive modeling in speech recognition systems. Unfortunately, the effects of Waibel's experiments are limited by his vocabulary sample, which contained a large number of monosyllabic words. I hope Waibel and others will replicate this study with a different and larger set of materials.

I recommend this book as a text and reference. Readers will benefit from a knowledge of speech basics, but no special knowledge of synthesis or recognition technology is necessary. Finally, I wish to recommend this volume especially to those who are currently looking at ways of improving a recognizer's performance through using better-known tools of computational linguistics, such as morphological analysis and parsing.

References

Lea, W. A. 1980 *Trends in Speech Recognition.* Prentice-Hall, Englewood Cliffs, NJ.

Joan Bachenko holds a Ph.D. in linguistics from New York University. Her research interests include prosody and speech synthesis, telegraphic sublanguages, and Deaf English. Bachenko's address is: AT&T Bell Laboratories, 3D462, Murray Hill, NJ 07974. E-mail:joan-b@allegra.att.com.

From Syntax to Semantics: Insights from Machine Translation

Erich Steiner, Paul Schmidt, and Cornelia Zelinsky-Wibbelt (eds.)
(Institut für Angewandte Informationsforschung, Saarbrücken)

London, U.K.: Pinter Publishers, 1988, vii + 262 pp.
(Communication in artificial intelligence series)
Hardbound, ISBN 0-86187-960-0, £29.50 and Norwood, NJ: Ablex Publishing Corp, 1988, vii + 262 pp.
Hardbound, ISBN 0-89391-526-2, $48.50 (institutional), $29.50 (personal)

Reviewed by
Harold Somers
UMIST

Keen observers of the world of machine translation have long awaited the first book-length publication on the Commission of the European Communities' MT project EUROTRA from the Saarbrücken-based German group; other readers may have been attracted by this book's title. Both, regrettably, risk disappointment.

The former will find this book nothing like the hoped-for detailed description of the world's biggest and best-funded MT project. Not only is the book explicitly "not some official report on EUROTRA work" (p. 1), but in several places it contradicts or argues against EUROTRA doctrine ("... the concept of transfer developed in this chapter is the opinion of the author, not necessarily the view underlying the project as a whole") (p. 161). However, it assumes in the readership either some prior knowledge of the fine details of the project and, especially, its formalisms and jargon, or else access to numerous Commission documents to which it makes frequent reference (particularly the "EUROTRA Reference Manual"), even though they are not in the public domain. A good (or bad) example of this occurs when we are asked (p. 13) to consult Arnold et al. 1985 (an internal report) for an explanation of the use in EUROTRA of the term "unification." What is the reader to make of this? All we get is a hint that this term might be being used in a nonstandard way, with no hint as to what it might mean here, and no reasonable chance of following up the reference. Similarly, the dual use of the term "translation" both in its everyday meaning (i.e. between natural languages) and to describe a relationship between representations (e.g. p. 21) is very misleading. Other examples include "euroversal" (e.g. pp. 5, 187), and reference to "the corpus" (p. 6), though we are not told which corpus, or of what it is a corpus.

This brings us to the second set of potentially disappointed readers: those expecting to read about syntax and semantics in an MT system. The recurring theme, inasmuch as there is one, is that a purely syntactic representation is not suitable for a 72-language-pair transfer MT system, and so some sort of "semantics" must be used. This