engineering standpoint) of an architecture that provides a clear separation of linguistic and nonlinguistic knowledge, since it is hard to see how shallow processing could be implemented without this separation.

Although SPAR is implemented as a complete natural language system, it unfortunately still has something of a "toy" flavor for two reasons. First of all, the data texts were written specifically for this project, although not by people who knew about SPAR. Although many of the phenomena in these texts undoubtedly occur in more realistic texts, the work would perhaps have been more convincing had Carter used texts written for other purposes. Naturally occurring texts often contain problematic constructions such as nominalizations, which present many interesting challenges for semantics and anaphor resolution in natural language systems (Dahl et al. 1987), but which don't occur in Carter's texts. The system also has a toy flavor because the end application, paraphrase, is less obviously useful than many other applications that might have been selected. There is no reason to think that these problems affect the fundamental soundness of the work, but they do tend to make it less interesting.

This work presents a very comprehensive implementation of the state of the art of reference resolution in natural language processing. However, one is left at the end with a frustrating sense that the whole process consists of exploiting a set of more or less unrelated heuristics, which in fact lead to very accurate reference resolution, but which don't seem to fit together into a general picture of a unified phenomenon. For example, Carter points out (using Sidner's terminology) that the discourse focus is preferred to intra-sentential candidates, but that intra-sentential candidates are preferred to potential discourse foci. It is only natural to wonder why these preferences (and others) should be the way they are, and whether they can be expected to fall out from more general principles. This is not specifically a criticism of Carter, but points out an unsatisfying aspect of much computational work in anaphora resolution. It is in fact at least partially the result of the clarity of his presentation that this issue emerges.

I found this book very stimulating, interesting, and clear. I would recommend it to anyone interested in reference resolution or computational pragmatics in general.

REFERENCES

Boguraev, B. K. 1979 *Automatic Resolution of Linguistic Ambiguity.* University of Cambridge Computer Laboratory, TR-11.

Dahl, Deborah A. 1985 *The Structure and Function of* one-*anaphora in English.* Indiana University Linguistics Club, Bloomington, Indiana.

Dahl, Deborah A.; Palmer, Martha S.; and Passonneau, Rebecca J. 1987 "Nominalizations in PUNDIT." *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics,* Stanford, CA.

Halliday, M. A. K. and Hasan, R. 1976 *Cohesion in English.* Longman, London.

Sidner, Candace L. 1979 *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse.* MIT Artificial Intelligence Laboratory, TR-537.

Webber, Bonnie L. 1978 *A Formal Approach to Discourse Anaphora.* Garland, New York.

Webber, Bonnie L. 1983 "So What Can We Talk About Now?" In Brady, M. and Berwick, R. (eds.), *Computational Models of Discourse.* MIT Press, Cambridge, MA.

Wilks, Yorick A. 1975 "A Preferential, Pattern-Seeking Semantics for Natural Language Inference." *Artificial Intelligence* 6:53–74.

Deborah Dahl received her Ph.D. in linguistics from the University of Minnesota in 1984. Her dissertation was on the interpretation of *one*-anaphora in discourse. She is currently a senior staff scientist in the natural language processing group at Unisys Paoli Research Center. Her group is developing PUNDIT, a large text processing system in Prolog. She designed and implemented PUNDIT's components for reference resolution and noun phrase semantics. She has also worked on the interpretation of indefinite noun phrases and the interaction of prosodic information with pronoun interpretation. Dahl's address is: Unisys Paoli Research Center, P.O. Box 517, Paoli, PA 19301. E-mail: dahl@prc.unisys.com

---

PHONOLOGICAL PARSING IN SPEECH RECOGNITION

Kenneth W. Church
(AT&T Bell Laboratories)

Reviewed by
Kimmo Koskenniemi
University of Helsinki

Church argues against the so-called "standard position" in speech recognition, i.e. the use of syntactic–semantic knowledge to disambiguate uncertain sounds in utterances. The argument proceeds by showing first that allophonic variation in speech is a source of useful information and not an obstacle for speech recognition. The claim sounds reasonable, but is not trivial because much of the work in the past was based on an opposite view. A part of this claim is that allophonic cues often indicate the location of boundaries (Nakatani's position).

Another important issue in the book is that syllable structure is very useful as a framework for describing allophonic variation. This is also a very reasonable claim from the linguistic point of view, but it is something that many of the leading current phonological theories fail to achieve.

The author assumes a constituency hypothesis where many allophonic and phonological processes share the same environments (e.g. foot-initial, foot-internal). This is done in a phrase-structure and chart-parsing framework. The

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)

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