

BOOK REVIEWS

INTERPRETING ANAPHORS IN NATURAL LANGUAGE TEXTS

David M. Carter

(SRI International, Cambridge)

Chichester, U.K.: Ellis Horwood, 1987, 292 pp.
(Ellis Horwood Series in Artificial Intelligence)
Hardbound, ISBN 0-470-21009-5, \$74.95

Reviewed by
Deborah Dahl
Unisys

This book explores an approach to text understanding, in particular anaphora resolution, that tries to limit the use of detailed domain knowledge and commonsense inference by exploiting general linguistic knowledge as much as possible. Carter proposes that, because natural language texts are relatively redundant and are constructed considerably, it should be possible in many cases to recover the interpretation of the text by using either linguistic or nonlinguistic techniques. Linguistic techniques are preferable because they are more general and less open-ended. Carter's approach is called "shallow processing." Since domain knowledge and reasoning tend to be expensive to implement, maintain, and use during processing, the shallow processing approach should be much more efficient and portable than an approach based heavily on domain-specific knowledge. If it can at the same time provide reasonable accuracy, then it should be very useful. This seems to be a very sensible approach in principle, and Carter demonstrates that it is quite effective in practice.

The shallow processing hypothesis is tested in a program called SPAR (Shallow Processing Anaphor Resolver), which was implemented as part of the author's University of Cambridge thesis. Shallow processing is presented as an engineering solution to the problem of dealing with the use of domain knowledge in text understanding, for certain applications, not as a psychological hypothesis. In the SPAR architecture, general linguistic techniques, such as focusing, are used first. Domain reasoning is only used if more than one candidate referent remains after the application of linguistic knowledge. Although this approach is tested specifically only for reference resolution, obviously it could be extended to other areas of natural language processing in which both linguistic and domain knowledge could be used, such as reasoning.

The book presents an excellent and very clear review of both current and older approaches to anaphora resolution, as well as a clear description of the SPAR system. For this reason, it would serve as a very good text for a seminar on

reference resolution as well as an extra reading for a class on knowledge representation.

One attractive aspect of the SPAR system is that it builds on previous work where appropriate, and extends it where required. In particular, it integrates the work of Boguraev (1979) in parsing, the work of Sidner (1979) in focusing, and the work of Wilks (1975) in preference semantics. Where Sidner's work, for example, is incomplete, as in the treatment of intrasentential anaphora, Carter presents a reasonable extension to handle the additional phenomena. One minor oversight in this work is that, in the treatment of *one*-anaphora, Carter fails to explore recent pragmatically oriented approaches, such as those discussed by Webber (1983) and Dahl (1984), who propose unified treatments of definite pronouns and *one*-anaphora. Instead, SPAR uses the older, and probably less effective, syntactic approach suggested by Webber (1978) and Halliday and Hasan (1976).

Another commendable aspect of this work is that Carter presents specific statistics on the accuracy of his system—93% of pronominal anaphors (out of 242) and 82% of nonpronominal anaphors (out of 80) are resolved correctly. Although these statistics go beyond what is usually reported, it would have been even more interesting to see a detailed breakdown of anaphor types and accuracy. It would also have been interesting to see statistics on the efficiency of the system, since the overall algorithm is quite complex. This high level of accuracy provides evidence that shallow processing is a promising approach. However, these statistics do raise the issue that we don't really know what level of accuracy in anaphora resolution is "good enough." In fact, the "good enough" level of accuracy may vary by application. Ninety-three percent may be accurate enough for some applications, such as machine translation, or message routing, but not for others, such as database update. Perhaps the relatively inexpensive shallow processing approach will turn out to be the method of choice for applications with lower accuracy requirements. A related issue that this work raises is how accurate we can expect to get—that is, how good is human performance on anaphora resolution, and how close is Carter's system to that level? These are questions that we simply don't know the answers to, and that await future research, Carter has done a valuable service in providing his statistics, but it is difficult to interpret them without having these bases of comparison and without having comparable statistics from other approaches.

In addition to supporting the shallow processing hypothesis, this work also supports the usefulness (at least from an

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)

Boston, MA: Kluwer Academic Publishers, 1987,
xv + 261 pp.

(The Kluwer International Series in Electrical
Engineering and Computer Science; VLSI, Computer
Architecture and Digital Signal Processing)

Hardbound, ISBN 0-89838-250-5, \$49.95

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