

## Anaphora Resolution

**Ruslan Mitkov**

(University of Wolverhampton)

London: Longman (Pearson Education),  
2002, xiii+220 pp; hardbound, ISBN  
0-582-32505-6, £49.99

*Reviewed by*

*Michael Strube*

*EML Research gGmbH*

Anaphora resolution is one of the themes in computational linguistics that seems to live forever. Although there has been a considerable amount of work in the past, researchers are still able to report on significant progress in this area—and it seems that the problems are far from being solved. The last monograph about anaphora resolution appeared more than 15 years ago with Carter's *Interpreting Anaphors in Natural Language Texts* (1987). Since then, anaphora resolution has taken an important role in applications such as summarization and open-domain question answering. These applications required a major shift in work on anaphora resolution: from theoretically motivated algorithms applicable to a few examples to robust systems dealing with unrestricted text. So it seems to be the right time for an update on monographs in the field.

Ruslan Mitkov "aims to fill an existing gap in the literature with ... an objective, comprehensive and up-to-date survey of the field," as he points out in the preface of his book, which, he notes, "also includes considerable discussion of [his] own research." The task of this review is to determine whether Mitkov provides in fact the long-awaited survey.

The book consists of nine chapters. The first one introduces the linguistic phenomenon of anaphora and important terminology. The computational treatment of anaphora is outlined in the second chapter, in which the first algorithmic aspects are given. Brief summaries of the centering model, the binding theory, the focus model, and the discourse representation theory in the third chapter are the theoretical background for the remainder of the book.

Chapter 4, "The Past: Work in the 1960s, 1970s and 1980s," describes the most important work from those decades. Only very early research relied primarily on heuristics. Mitkov describes later research in that area as "theoretically-oriented and rather ambitious in terms of the types of anaphora handled." It extensively exploited linguistic knowledge (syntax, semantics, discourse) and nonlinguistic domain and world knowledge and did not focus on empirical issues. If an evaluation was provided, it was a small-scale one. The syntax-based pronoun resolution algorithm by Hobbs (1978), however, is still used as a hard-to-beat benchmark today.

Mitkov's focus becomes narrower in chapter 5, "The Present: Knowledge-Poor and Corpus-Based Approaches in the 1990s and Beyond." In this chapter he sets up the context that forms the foundation for his own research. He gives an extensive description of the important heuristic algorithm by Lappin and Leass (1994) and describes so-called knowledge-poor approaches in detail. In contrast to the methods of Hobbs (1978) and Lappin and Leass (1994), these approaches do not rely on a complete syntactic analysis. They do not need domain and world knowledge either. Instead, POS taggers and NP chunkers or partial parsers are used, which allow pronoun resolu-

tion to be performed robustly on unrestricted text without sacrificing performance. Mitkov motivates this approach justly with regard to applications like summarization, open-domain question answering, and so on. He also discusses briefly a few machine-learning and statistical approaches (e.g., Soon, Ng, and Lim 2001).

Chapter 6, “The Role of Corpora in Anaphora Resolution,” serves as an interlude before Mitkov finally presents his own approach in chapter 7, “An Approach in Focus: Mitkov’s Robust Knowledge-Poor Algorithm.” He describes several stages of his work on pronoun resolution. He begins with his “original approach,” which consists of a set of heuristics in the spirit of the knowledge-poor approaches described in chapter 5. Mitkov then shows how his approach can be applied to languages other than English (Polish, Arabic, French) and how it benefits from being applied to parallel corpora (English, French). Whereas the original approach expects manually corrected input, a reimplemented version of the approach called MARS (Mitkov’s anaphora resolution system) gets its input from a chain of preprocessing modules. Hence this version is fully automatic and does not depend on manual intervention. Chapter 8, “Evaluation in Anaphora Resolution,” introduces Mitkov’s new evaluation metric and reports results for the original approach, MARS, several baselines, and other systems using this metric. Mitkov concludes the book with a chapter on outstanding issues.

The book has several strengths, but unfortunately also weaknesses. On the strong side there is a well-chosen, comprehensive, and easily accessible overview of important work in the past. The first part of the book can be recommended to anyone new to the field. This part of the book not only gives an overview of past work, but also presents many points that may help even seasoned practitioners to improve their anaphora resolution algorithms. For example, section 2.2.1 gives some information on the identification of anaphors that at least this reviewer did not know before. The same is true for section 2.2.3 and chapter 4, which describe factors used for anaphora resolution. Some of these factors were brought to this reviewer’s attention only by the book.

However, Mitkov’s overview of present work seems to be biased. He focuses almost exclusively on the knowledge-poor approaches that had their heyday in the mid- and late 1990s. Since then, the advantages of machine-learning approaches have led to a decline of the knowledge-poor approaches. Machine-learning approaches prevailed because they performed as well as knowledge-poor approaches but were more robust, more domain-independent, and—in contrast to the knowledge-poor approaches— theoretically more sound. Whereas heuristics used by the knowledge-poor approaches need to be manually tuned and optimized, machine-learning algorithms tune themselves during training. So in a comprehensive and up-to-date survey of the field, one would expect a more thorough discussion of machine-learning approaches to anaphora resolution.

The same criticism can be applied to Mitkov’s chapter on evaluation. In section 8.2, he briefly points out inconsistencies in earlier definitions of *precision* and *recall*. He uses this as motivation for his own proposal, which takes up most of this chapter. Only once, however, does he mention the widely used model-theoretic evaluation metric developed by Vilain et al. (1995) for the coreference task in the MUC conferences. It would have been very helpful if Mitkov had explained this inherently difficult-to-understand but very important evaluation metric. The benefit of Mitkov’s chapter on evaluation is that he makes it clear that there are vast differences between the evaluation of anaphora resolution algorithms (expecting perfect input) and anaphora resolution systems (depending on a chain of preprocessing modules yielding error-prone input).

In addition, it is necessary to point out that the title of the book is slightly misleading. Though the book is called *Anaphora Resolution*, the work described mainly

concerns pronoun resolution. There are only a few passages dedicated to anaphora resolution beyond pronouns, such as the resolution of definite noun phrases (e.g., section 5.6, which briefly describes the approach by Vieira and Poesio [2000]). Factors used for this task are discussed only briefly within the context of machine-learning approaches (e.g., Soon, Ng, and Lim 2001). Researchers looking for information about anaphora resolution in general will not get much information from Mitkov's book.

Since Mitkov is a devotee of the knowledge-poor approaches, which do not have much theoretical underpinning, he does not provide a systematic synthesis at the end of the book, neither theoretically nor empirically. He does not say *why* certain approaches work better than other approaches. He does not provide an overview of similarities and differences in the approaches discussed in terms of the factors they use.

On a final note, I got the impression that this is not the book about anaphora resolution that could fill the large gap since Carter's monograph. Mitkov's book (especially the first chapters) can be recommended to anyone new to the field or researchers interested in knowledge-poor approaches for pronoun resolution. If, however, you are interested in either theoretical or machine-learning approaches or if you are looking for information about anaphora beyond pronouns, you will have to wait for the next book on that topic.

#### References

- Carter, David. 1987. *Interpreting Anaphors in Natural Language Texts*. Ellis Horwood, Chichester, England.
- Hobbs, Jerry R. 1978. Resolving pronominal references. *Lingua*, 44:311–338.
- Lappin, Shalom and Herbert J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–561.
- Soon, Wee Meng, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544.
- Vieria, Renata and Massimo Poesio. 2000. An empirically-based system for processing definite descriptions. *Computational Linguistics*, 26(4):539–593.
- Vilain, Marc, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A model-theoretic coreference scoring scheme. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 45–52, San Mateo, CA. Morgan Kaufmann.

*Michael Strube* is the leader of the Natural Language Processing Group at the EML Research in Heidelberg, Germany. His research interests include discourse and dialogue processing with a particular focus on anaphora resolution. E-mail: [strube@eml-research.de](mailto:strube@eml-research.de); URL: <http://www.eml-research.de/english/homes/strube>.