Studies in Natural Language Processing

Challenges in natural language processing
Studies in Natural Language Processing
Branimir K. Boguraev, Editor

This series publishes monographs, texts, and edited volumes within the interdisciplinary field of computational linguistics. It represents the range of topics of concern to the scholars working in this increasingly important field, whether their background is in formal linguistics, psycholinguistics, cognitive psychology, or artificial intelligence.

Also in this series:

*Memory and context for language interpretation* by Hiyan Alshawi
*The linguistic basis of text generation* by Laurence Danlos
*Natural language parsing* edited by David R. Dowty, Lauri Karttunen, and Arnold Zwicky
*Relational models of the lexicon* by Martha Walton Evens
*Computational linguistics* by Ralph Grishman
*Semantic interpretation and the resolution of ambiguity* by Graeme Hirst
*Reference and computation* by Amichai Kronfeld
*Machine translation* edited by Sergei Nirenburg
*Semantic processing for finite domains* by Martha Stone Palmer
*Systemic text generation as problem solving* by Terry Patten
Challenges in natural language processing

Edited by
MADELEINE BATES
and
RALPH M. WEISCHDEL

BBN Systems and Technologies
Cambridge, MA
Contents

Preface \hspace{100pt} page vii

Symposium participants \hspace{100pt} x

Part I. Challenging problems
1 Critical challenges for natural language processing 3
   Madeleine Bates, Robert J. Bobrow, and Ralph M. Weischedel

Part II. Building a lexicon
2 The contribution of lexicography 37
   B. T. Sue Atkins
3 The contribution of linguistics 76
   Beth Levin
4 The contribution of computational lexicography 99
   Branimir K. Boguraev

Part III. Semantics and knowledge representation
5 Events, situations, and adverbs 135
   Robert C. Moore
6 Natural language, knowledge representation, and logical form 146
   James F. Allen

Part IV. Discourse
7 Getting and keeping the center of attention 179
   Rebecca J. Passonneau
8 Surface structure, intonation, and discourse meaning 228
   Mark Steedman

Part V. Spoken language systems
9 Prosody, intonation, and speech technology 257
   Janet Pierrehumbert
## Contents

### Part VI. Conclusion

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>The future of computational linguistics</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>Madeleine Bates and Ralph M. Weischedel</td>
<td></td>
</tr>
</tbody>
</table>

**Author index**

Page 289

**Subject index**

Page 292
Preface

Throughout most of the 1960s, 1970s, and 1980s, computational linguistics enjoyed an excellent reputation. A sense of the promise of the work to society was prevalent, and high expectations were justified by solid, steady progress in research.

Nevertheless, by the close of the 1980s, many people openly expressed doubt about progress in the field. Are the problems of human language too hard to be solved even in the next eighty years? What measure(s) (other than the number of papers published) show significant progress in the last ten years? Where is the technology successfully deployed in the military, with revolutionary impact? In what directions should the field move, to ensure progress and avoid stagnation?

The symposium

A symposium on Future Directions in Natural Language Processing was held at Bolt Beranek and Newman, Inc. (BBN), in Cambridge, Massachusetts, from November 29, 1989, to December 1, 1989.

The symposium, sponsored by BBN’s Science Development Program, brought together top researchers in a variety of areas to discuss the most significant problems and challenges that will face the field of computational linguistics in the next two to ten years. Speakers were encouraged to present both recently completed and speculative work, and to focus on topics that will have the most impact on the field in the coming decade. They were asked to reconsider unsolved problems of long standing as well as to present new opportunities. The purpose was to contribute to long-range planning by funding agencies, research groups, academic institutions, graduate students, and others interested in computational linguistics.

The thirty-six symposium attendees, who are listed following this preface, were all invited participants. They included speakers, respondents to papers, and people from government, industry, and academia. Though varying greatly in age and in backgrounds, they all shared a strong interest in computational linguistics.

The format of the meeting provided for one-and-a-half-hour presentations, much longer than at most conferences, in order to give speakers sufficient time to present ideas completely. The schedule also allowed plenty of time for lively
Preface

discussions. A brief review of the symposium was published by Mark Maybury in the Summer 1990 issue of AI Magazine.

The papers and the book

We have taken the liberty of arranging the chapters in this book in a slightly different order than they were presented at the symposium. This allows our own paper, “Critical Challenges for NLP,” to come first, both to set the tone for the subsequent papers and to provide a number of definitions and examples that may be helpful to readers who are not fully familiar with this field.

The three “traditional” areas of NLP – syntax, semantics, and pragmatics – all focus on language that is made up of words (or entities very similar to words). It is curious that the study of words themselves has not traditionally been a major area of interest in computational linguistics. Perhaps this is because it has always been assumed that each NL system has its own, usually small (under 5,000 words) lexicon that has been developed by hand for the application under study.

One of the key insights of the BBN symposium is that more attention must be paid to the development of large, sharable lexicons that are produced specifically for use by NL systems. Traditional dictionaries have been produced for use in book form by humans, but this is completely inadequate for computational applications.

The three papers that form Part II, Building the lexicon, address these issues. Sue Atkins (Oxford University Press), whose talk was originally titled “Beware the lexicon” but appears here as “The contribution of lexicography,” gives a lexicographer’s view of a proposed structure for lexical entries. Beth Levin (Northwestern University) provides an analysis of verbs, specifically verbs of making sound, that allows for the handling of new uses of known words. Bra-nimir Boguraev (IBM Yorktown Heights) discusses the use of machine-readable sources to construct thesauri and other knowledge bases.

The areas of semantics and knowledge representation remain closely coupled and of great import. Robert Moore (SRI International) presents a study of the adverb, a part of speech that has been largely ignored in the past. James Allen (University of Rochester) provides an excellent discussion of the use of ambiguity and a related concept, vagueness, in logical form.

The subject of discourse has been receiving increased attention in the past few years. Rebecca Passonneau (Paramax) addresses this important issue. Mark Steedman (University of Pennsylvania) originally titled his presentation “Prosody, parsing, and discourse,” and thus it is included in the discourse section, although its title here is “Surface structure, intonation, and discourse meaning.”

It has long been the goal of many researchers to integrate speech processing techniques (such as speech recognition) with natural language processing to produce systems capable of understanding spoken language. Janet Pierrehumbert
(Northwestern University) argues that a rich model of prosodic structure is necessary to understand how prosody and intonation convey information about attention and intention.

The book concludes with the results of a lengthy and lively discussion on the future of computational linguistics. This discussion attempted to enumerate some of the most important problems to be tackled in the next few years and to describe what will be needed to solve them.

The papers that were presented at the BBN symposium represent not only cutting-edge research on familiar topics but also new directions for the field. This is indicated by the presence of such topics as the lexicon and speech (and the absence of such topics as parsers and grammars), which would not have been the case in a similar meeting held even five years ago.

The purpose of this volume is to convey that excitement, to suggest concrete ideas that the contributing authors think most promising for future research, and to encourage the readers to pursue ideas for a new generation of NLP technology.

Madeleine Bates
Ralph M. Weischedel
Symposium participants

JAMES F. ALLEN
University of Rochester (s) (r)

B. T. SUE ATKINS
Oxford University Press (s)

MADEleine BATES
BBN Systems & Technologies (s)

ROBERT J. BOBROW
BBN Systems & Technologies (s) (r)

BRANIMIR K. BOGURAEV
IBM Research (s)

Yen-Lu CHOW
BBN Systems & Technologies

STEPHEN DENNIS
Department of Defense

JOYCE FRIEDMAN
Boston University

JANE GRIMSHAW
Brandeis University

HENRY HAMBURGER
George Mason University

ROBERT J. P. INGRIA
BBN Systems & Technologies

BORIS KATZ
MIT AI Laboratory

BETH LEVIN
Northwestern University

MARK MAYBURY
RADC

ROBERT MERCER
IBM Research (r)

CHRISTENE MONTGOMERY
LIS (r)

ROBERT C. MOORE
SRI International (s)

REBECCA J. PASSONEAU
Paramax (s)

JANET PIERREHUMBERT
Northwestern University (s)

JAMES PUSTEJOVKY
Brandeis University (r)

RICHARD SCHWARTZ
BBN Systems & Technologies (s)

STEPHANIE SENEFF
MIT

VARDA SHAKED
BBN Systems & Technologies

CANDY SIDNER
Digital Equipment Corporation (r)

NORM SONDEHIMER
GE Research (r)

MARK STEEDMAN
University of Pennsylvania (s)

MIKE THOMAS
KSC, Inc.

DONALD WALKER
Bellcore (r)

(s) indicates a speaker.
(r) indicates a respondent.

X
Symposium participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharon Walter</td>
<td>RALPH M. WEISCHEDEL</td>
</tr>
<tr>
<td>RADC</td>
<td>BBN Systems &amp; Technologies (s)</td>
</tr>
<tr>
<td>Rand Waltzman</td>
<td>DOUGLAS WHITE</td>
</tr>
<tr>
<td>DARPA/ISTO</td>
<td>RADC</td>
</tr>
<tr>
<td>Charles L. Wayne</td>
<td>COLIN WIGHTMAN</td>
</tr>
<tr>
<td>DARPA/ISTO</td>
<td>Boston University</td>
</tr>
<tr>
<td>Bonnie Lynn Webber</td>
<td>WILLIAM WOODS</td>
</tr>
<tr>
<td>University of Pennsylvania</td>
<td>Harvard University (r)</td>
</tr>
</tbody>
</table>