

Second
Conference
on
Applied
Natural Language
Processing

Association for
Computational Linguistics

Proceedings of the Conference

9–12 February 1988
Austin-Marrriott at the Capitol
Austin, Texas, USA

Published by the Association for Computational Linguistics

© 1988, Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Donald E. Walker (ACL)
Bell Communications Research
445 South Street MRE 2A379
Morristown, NJ 07960-1961, USA

PREFACE

This *Second Conference on Applied Natural Language Processing* focuses on “the application of natural language processing techniques to real world problems” and represents many of the advances that have been made since the Association for Computational Linguistics organized its first such conference four years ago in Santa Monica, California.

In addition to presentations of thirty-two papers, a panel discussion on “Natural Language Interfaces: Present and Future,” and six tutorials, the conference includes a variety of exhibits and demonstrations. The papers, the introduction to the panel session, and abstracts of the tutorials are included in this volume.

Such a meeting can be a success only through the efforts of a large number of people. We are first indebted to the authors of the 106 papers that were received, representing over a dozen countries. While the vast majority of these papers reported on interesting and important work, fewer than a third of the papers could be included in the program. For having to make the difficult decisions, we are also indebted to the members of the Program Committee:

Madeleine Bates	<i>BBN Laboratories</i>
Tim Finin	<i>Unisys</i>
Ralph Grishman	<i>New York University</i>
Carole Hafner	<i>Northeastern University</i>
George Heidorn	<i>IBM Corporation</i>
Paul Martin	<i>SRI International</i>
Graeme Ritchie	<i>University of Edinburgh</i>
Harry Tennant	<i>Texas Instruments</i>

In addition, much credit and many thanks go to Martha Palmer (Unisys), who organized the tutorials; to Jonathan Slocum (MCC), who handled local arrangements; to Kent Wittenburg and Rich Cohen (MCC), for co-ordinating the exhibits and demonstrations; to Brenda Nashawaty (Artificial Intelligence Corporation), who handled publicity; to Gary Hendrix (Symantec), for providing the facilities within *Q&A* for membership and conference registration information; to AT&T Bell Laboratories, for providing secretarial support and for hosting the program committee meeting; to MCC, for sponsoring a reception during the conference; and to Mike Dunkle (Austin Mariott), for many ways of helping to prepare for the meeting.

Finally, special thanks go to Norm Sondheimer (General Chair, USC/Information Sciences Institute) and Don Walker (ACL Secretary-Treasurer, Bell Communications Research), for attending to an array of issues that arose, many of which most of us never knew existed.

Bruce Ballard, *AT&T Bell Laboratories*
Chair, Program Committee

CONFERENCE PROGRAM

TUESDAY, 9 FEBRUARY

Joe C. Thompson Conference Center, University of Texas at Austin

TUTORIALS

8:30–12:30 MORNING SESSIONS

Introduction to Natural Language Processing

James Allen (University of Rochester)

Machine-Readable Dictionaries: A Computational Linguistics Perspective

Bran Boguraev (Cambridge University) and Beth Levin (Northwestern University)

Spoken Language Systems: Past, Present, and Future

Salim Roucos (BBN Laboratories, Inc.)

1:30–5:30 AFTERNOON SESSIONS

The Technology of Natural Language Interfaces

Carole Hafner (Northeastern University)

The Role of Logic in Representing Meaning and Knowledge

Bob Moore (SRI International)

Machine Translation

Sergei Nirenburg (Carnegie-Mellon University)

RECEPTION

7:00–9:00 Austin Marriott at the Capitol

WEDNESDAY, 10 FEBRUARY

Austin Marriott at the Capital

9:00–9:15 Opening remarks and announcements

SESSION 1: SYSTEMS

9:15–9:40 *The Multimedia Articulation of Answers in a Natural Language Database Query System*
Susan E. Brennan (Stanford University & Hewlett Packard)

9:40–10:05 *A News Story Categorization System*
Philip J. Hayes, Laura E. Knecht and Monica J. Cellio (Carnegie Group)

10:05–10:30 *An Architecture for Anaphora Resolution*
Elaine Rich and Susann LuperFoy (MCC)

SESSION 2: GENERATION

11:00–11:25 *The SEMSYN Generation System: Ingredients, Applications, Prospects*
Dietmar Roesner (Universitaet Stuttgart)

11:25–11:50 *Two Simple Prediction Algorithms to Facilitate Text Production*
Lois Boggess (Mississippi State University)

11:50–12:15 *From Water to Wine: Generating Natural Language Text from Today's Applications Programs*
David D. McDonald (Brattle Research Corporation) and Marie M. Meteer (BBN Laboratories)

LUNCHEON

12:15–2:00 *Guest Speaker: Grant Dove*
Chairman and CEO of MCC. Prior to joining MCC in July 1987, Mr. Dove had been with Texas Instruments for 28 years, having served as Executive Vice President since 1982.

SESSION 3: SYNTAX AND SEMANTICS

- 2:00–2:25 *Improved Portability and Parsing Through Interactive Acquisition of Semantic Information*
Francois-Michel Lang and Lynette Hirschman (UNISYS)
- 2:25–2:50 *Handling Scope Ambiguities in English*
Sven Hurum (University of Alberta)
- 2:50–3:15 *Responding to Semantically Ill-Formed Input*
Ralph Grishman and Ping Peng (New York University)
- Evaluation of a Parallel Chart Parser*
Ralph Grishman and Mahesh Chitrao (New York University)

SESSION 4: MORPHOLOGY AND THE LEXICON

- 3:45–4:10 *Triphone Analysis: A Combined Method for the Correction of Orthographical and Typographical Errors*
Brigitte van Berkel (TNO Institute of Applied Computer Science) and
Koenraad DeSmedt (University of Nijmegen)
- 4:10–4:35 *Creating and Querying Hierarchical Lexical Data Bases*
Mary S. Neff, Roy J. Byrd, and Omneya A. Rizk (IBM Watson Research Center)
- 4:35–5:00 *Cn yur cmputr raed ths?*
Linda G. Means (General Motors)
- 5:00–5:25 *Building a Large Thesaurus for Information Retrieval*
Edward A. Fox, J. Terry Nutter (Virginia Tech), Thomas Ahlswede, Martha Evens
(Illinois Institute of Technology), and Judith Markowitz (Navistar International)

RECEPTION

- 6:30–9:00 Microelectronics and Computer Technology Corporation (MCC)

THURSDAY, 11 FEBRUARY Austin Marriott at the Capital

SESSION 5: SYSTEMS

- 8:30–8:55 *Application-Specific Issues in Natural Language Interface Development for a Diagnostic Expert System*
Karen L. Ryan, Rebecca Root, and Duane Olawsky (Honeywell)
- 8:55–9:20 *The MULTIVOC Text-to-Speech System*
Olivier Emorine and Pierre Martin (Cap Sogeti Innovation)
- 9:20–9:45 *Structure from Anarchy: Meta Level Representation of Expert System Propositions for Natural Language Interfaces*
Galina Datskovsky Moerdler (Columbia University)

SESSION 6: TEXT PROCESSING

- 10:15–10:40 *Integrating Top-Down and Bottom-Up Strategies in a Text Processing System*
Lisa F. Rau and Paul S. Jacobs (General Electric)
- 10:40–11:05 *A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text*
Kenneth W. Church (AT&T Bell Laboratories)
- 11:05–11:30 *A Tool for Investigating the Synonymy Relation in a Sense Disambiguated Thesaurus*
Martin S. Chodorow, Yael Ravin (IBM Watson Research Center) and Howard E. Sachar
(IBM Data Systems Division)
- 11:30–11:55 *Dictionary Text Entries as a Source of Knowledge for Syntactic and Other Disambiguations*
Karen Jensen and Jean-Louis Binot (IBM Watson Research Center)

LUNCHEON

12:00–1:45 Guest Speaker: **Donald E. Walker**
Manager of Artificial Intelligence and Information Science Research at Bell Communications Research, and Secretary-Treasurer of ACL and IJCAI

SESSION 7: MACHINE TRANSLATION

1:45–2:10 *EUROTRA: Practical Experience with a Multilingual Machine Translation System under Development*
Giovanni B. Varile and Peter Lau (Commission of the European Communities)

2:10–2:35 *Valency and MT: Recent Developments in the METAL System*
Rudi Gebruers (Katholieke Universiteit Leuven)

3:00–5:00 **PANEL: Natural Language Interfaces: Present and Future**
Moderator: Norman Sondheimer (USC/Information Sciences Institute)
Panelists: Robert J. Bobrow (BBN Laboratories), Developer of *RUS*
Jerrold Ginsparg (Natural Language Inc.), Developer of *DataTalker*
Larry Harris (Artificial Intelligence Corporation), Developer of *Intellect*
Gary G. Hendrix (Symantec), Developer of *Q&A*
Steve Klein (Singular Solutions Engineering) Co-Developer of Lotus *HAL*

RECEPTION

5:00–6:00 Austin Marriott at the Capitol

FRIDAY, 12 FEBRUARY Austin Marriott at the Capital

SESSION 8: SYSTEMS

8:30–8:55 *Automatically Generating Natural Language Reports in an Office Environment*
Jugal Kalita and Sunil Shende (University of Pennsylvania)

8:55–9:20 *Luke: An Experiment in the Early Integration of Natural Language Processing*
David A. Wroblewski and Elaine A. Rich (MCC)

9:20–9:45 *The Experience of Developing a Large-Scale Natural Language Text Processing System: CRITIQUE*
Stephen D. Richardson and Lisa C. Braden-Harder (IBM Watson Research Center)

SESSION 9: MORPHOLOGY AND THE LEXICON

10:15–10:40 *Computational Techniques for Improved Name Search*
Beatrice T. Oshika (Sparta), Filip Machi (UC Berkeley), Bruce Evans (TRW), and Janet Tom (Systems Development Corporation)

10:40–11:05 *The TICC: Parsing Interesting Text*
David Allport (University of Sussex)

11:05–11:30 *Finding Clauses in Unrestricted Text by Finitary and Stochastic Methods*
Eva Ejerhed (University of Umea)

11:30–11:55 *Morphological Processing in the Nabu System*
Jonathan Slocum (MCC)

SESSION 10: SYNTAX AND SEMANTICS

1:30–1:55 *Localizing Expression of Ambiguity*
John Bear and Jerry R. Hobbs (SRI International)

1:55–2:20 *Combinatorial Disambiguation*
Paula S. Newman (IBM Los Angeles Scientific Center)

2:20–2:45 *Canonical Representation in NLP System Design: A Critical Evaluation*
Kent Wittenburg and Jim Barnett (MCC)

TUTORIAL ABSTRACTS

Introduction to Natural Language Processing

James Allen, *University of Rochester*

This tutorial will cover the basic concepts underlying the construction of natural language processing systems. These include basic parsing techniques, semantic interpretation and the representation of sentence meaning, as well as knowledge representation and techniques for understanding natural language in context. In particular, the topics to be addressed in detail will include augmented transition networks (ATNs), augmented context-free grammars, the representation of lexical meaning, especially looking at case-grammar based representations, and the interpretation of pronouns and ellipsis. In addition, there will be an overview of knowledge representation, including semantic networks, frame-based systems, and logic, and the use of general world knowledge in language understanding, including scripts and plans.

Given the large range of issues and techniques, an emphasis will be placed on those aspects relevant to existing practical natural language systems, such as interfaces to database systems. The remaining issues will be more quickly surveyed to give the attendee an idea of what techniques will become important in the next generation of natural language systems. The lecture notes will include an extensive bibliography of work in each area.

Machine-Readable Dictionaries: A Computational Linguistics Perspective

Branimir Boguraev, *Cambridge University*
and

Beth Levin, *Northwestern University*

The lexical information contained explicitly and implicitly in machine-readable dictionaries (MRDs) can support a wide range of activities in computational linguistics, both of theoretical interest and of practical importance. This tutorial falls into two parts. The first part will focus on some characteristics of raw lexical data in electronic sources, which make MRDs particularly relevant to natural language processing applications. The second part will discuss how theoretical linguistic research into the lexicon can enhance the contribution of MRDs to applied computational linguistics.

The first half will discuss issues concerning the placement of rich lexical resources on-line; raise questions related to the suitability, and ultimately the utility, of MRDs for automatic natural language processing; outline a methodology aimed at extracting maximally usable subsets of the dictionary with minimal introduction of errors; and present ways in which specific use can be made of the lexical data for the construction of practical language processing systems with substantial coverage.

The second half of the tutorial will review current theoretical linguistic research on the lexicon, emphasizing proposals concerning the nature of lexical representation and lexical organization. This overview will provide the context for an examination of how the results of this research can be brought to bear on the problem of extracting syntactic and semantic information encoded in dictionary entries, but not overtly signaled to the dictionary user.

Spoken Language Systems: Past, Present, and Future

Salim Roucos, *BBN Laboratories, Inc.*

This tutorial will present the issues in developing spoken language systems for natural speech communication between a person and a machine. In particular, the performance of complex tasks using large vocabularies and unrestricted sentence structures will be examined. The first Advanced Research Projects Agency (ARPA) Speech Understanding Research project during the seventies will be reviewed, and then the current state-of-the-art in continuous speech recognition and natural language processing will be described. Finally, the types of spoken language systems' capabilities expected to be developed during the next two to three years will be presented.

The technical issues that will be covered include acoustic-phonetic modeling, syntax, semantics, plan recognition and discourse, and the issues for integrating these knowledge sources for speech understanding. In addition, computational requirements for real-time understanding, and performance evaluation methodology will be described. Some of the human factors of speech understanding in the context of performing interactive tasks using an integrated interface will also be discussed.

The Technology of Natural Language Interfaces

Carole D. Hafner, *Northeastern University*

This tutorial will describe the development of natural language processing from a research topic into a commercial technology. This will include a description of some key research projects of the 1970's and early 1980's which developed methods for building natural language query interfaces, initially restricted to just one database, and later made "transportable" to many different applications. The further development of this technology into commercial software products will be discussed and illustrated by a survey of several current products, including both micro-computer NL systems and those offered on higher-performance machines. The qualities a user should look for in a NL interface will be considered, both in terms of linguistic capabilities and general ease of use. Finally, some of the remaining "hard problems" that current technology has not yet solved in a satisfactory way will be discussed.

The Role of Logic in Representing Meaning and Knowledge

Robert C. Moore, *SRI International*

This tutorial will survey the use of logic to represent the meaning of utterances and the extra-linguistic knowledge needed to produce and interpret utterances in natural-language processing systems. Problems to be discussed in meaning representation include quantification, propositional attitudes, comparatives, mass terms and plurals, tense and aspect, and event sentences and adverbials. Logic-based methods (unification) for systematic specification of the correspondence between syntax and semantics in natural language processing systems will also be touched on. In the discussion of the representation of extra-linguistic knowledge, special attention will be devoted to the role played by knowledge of speakers' and hearers' mental states (particularly their knowledge and beliefs) in the generation and interpretation of utterances and logical formalisms for representing and reasoning about knowledge of those states.

Machine Translation

Sergei Nirenburg, *Carnegie Mellon University*

The central problems faced by a Machine Translation (MT) research project are 1) the design and implementation of automatic natural language analyzers and generators that manipulate morphological, syntactic, semantic and pragmatic knowledge; and 2) the design, acquisition and maintenance of dictionaries and grammars. Since a short-term goal (or even medium term goal) of building a system that performs fully automated machine translation of unconstrained text is not feasible, an MT project must carefully constrain its objectives.

This tutorial will describe the knowledge and processing requirements for an MT system. It will present and analyze the set of design choices for MT projects including distinguishing features such as long-term/short-term, academic/commercial, fully/partially automated, direct/transfer/interlingua, pre-/post-/interactive editing. The knowledge acquisition needs of an MT system, with an emphasis on interactive knowledge acquisition tools that facilitate the task of compiling the various dictionaries for an MT system will be discussed. In addition, expectations, possibilities and prospects for immediate application of machine translation technology will be considered. Finally, a brief survey of MT research and development work around the world will be presented.

TABLE OF CONTENTS

SESSION 1: SYSTEMS

<i>The Multimedia Articulation of Answers in a Natural Language Database Query System</i> Susan E. Brennan	1
<i>A News Story Categorization System</i> Philip J. Hayes, Laura E. Knecht, and Monica J. Cellio	9
<i>An Architecture for Anaphora Resolution</i> Elaine Rich and Susann LuperFoy	18

SESSION 2: GENERATION

<i>The SEMSYN Generation System: Ingredients, Applications, Prospects</i> Dietmar Roesner	25
<i>Two Simple Prediction Algorithms to Facilitate Text Production</i> Lois Boggess	33
<i>From Water to Wine: Generating Natural Language Text from Today's Applications Programs</i> David D. McDonald and Marie M. Meteer	41

SESSION 3: SYNTAX AND SEMANTICS

<i>Improved Portability and Parsing Through Interactive Acquisition of Semantic Information</i> Francois-Michel Lang and Lynette Hirschman	49
<i>Handling Scope Ambiguities in English</i> Sven Hurum	58
<i>Responding to Semantically Ill-Formed Input</i> Ralph Grishman and Ping Peng	66
<i>Evaluation of a Parallel Chart Parser</i> Ralph Grishman and Mahesh Chitrao	71

SESSION 4: MORPHOLOGY AND THE LEXICON

<i>Triphone Analysis: A Combined Method for the Correction of Orthographical and Typographical Errors</i> Brigitte van Berkel and Koenraad DeSmedt	77
<i>Creating and Querying Hierarchical Lexical Data Bases</i> Mary S. Neff, Roy J. Byrd, and Omneya A. Rizk	84
<i>Cn yur cmputr raed ths?</i> Linda G. Means	93
<i>Building a Large Thesaurus for Information Retrieval</i> Edward A. Fox, J. Terry Nutter, Thomas Ahlswede, Martha Evens, and Judith Markowitz	101

SESSION 5: SYSTEMS

<i>Application-Specific Issues in Natural Language Interface Development for a Diagnostic Expert System</i> Karen L. Ryan, Rebecca Root, and Duane Olawsky	109
<i>The MULTIVOC Text-to-Speech System</i> Olivier Emorine and Pierre Martin	115
<i>Structure from Anarchy: Meta Level Representation of Expert System Propositions for Natural Language Interfaces</i> Galina Datskovsky Moerdler	121

SESSION 6: TEXT PROCESSING

<i>Integrating Top-Down and Bottom-Up Strategies in a Text Processing System</i> Lisa F. Rau and Paul S. Jacobs	129
<i>A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text</i> Kenneth W. Church	136
<i>A Tool for Investigating the Synonymy Relation in a Sense Disambiguated Thesaurus</i> Martin S. Chodorow, Yael Ravin, and Howard E. Sachar	144
<i>Dictionary Text Entries as a Source of Knowledge for Syntactic and Other Disambiguations</i> Karen Jensen and Jean-Louis Binot	152

SESSION 7: MACHINE TRANSLATION

<i>EUROTRA: Practical Experience with a Multilingual Machine Translation System under Development</i> Giovanni B. Varile and Peter Lau	160
<i>Valency and MT: Recent Developments in the METAL System</i> Rudi Gebruers	168

PANEL STATEMENT

<i>Natural Language Interfaces: Present and Future</i> Norman Sondheimer	176
---	-----

SESSION 8: SYSTEMS

<i>Automatically Generating Natural Language Reports in an Office Environment</i> Jugal Kalita and Sunil Shende	178
<i>Luke: An Experiment in the Early Integration of Natural Language Processing</i> David A. Wroblewski and Elaine A. Rich	186
<i>The Experience of Developing a Large-Scale Natural Language Text Processing System: CRITIQUE</i> Stephen D. Richardson and Lisa C. Braden-Harder	195

SESSION 9: MORPHOLOGY AND THE LEXICON

<i>Computational Techniques for Improved Name Search</i> Beatrice T. Oshika, Filip Machi, Bruce Evans, and Janet Tom	203
<i>The TICC: Parsing Interesting Text</i> David Allport	211
<i>Finding Clauses in Unrestricted Text by Finitary and Stochastic Methods</i> Eva Ejerhed	219
<i>Morphological Processing in the Nabu System</i> Jonathan Slocum	228

SESSION 10: SYNTAX AND SEMANTICS

<i>Localizing Expression of Ambiguity</i> John Bear and Jerry R. Hobbs	235
<i>Combinatorial Disambiguation</i> Paula S. Newman	243
<i>Canonical Representation in NLP System Design: A Critical Evaluation</i> Kent Wittenburg and Jim Barnett	253

AUTHOR INDEX

Thomas Ahlswede	101	Peter Lau	160
David Allport	211	Susann LuperFoy	18
Jim Barnett	253	Filip Machi	203
John Bear	235	Judith Markowitz	101
Jean-Louis Binot	152	Pierre Martin	115
Lois Boggess	33	David D. McDonald	41
Lisa C. Braden-Harder	195	Linda G. Means	93
Susan E. Brennan	1	Marie M. Meteer	41
Roy J. Byrd	84	Galina Datskovsky Moerdler	121
Monica J. Cellio	9	Mary S. Neff	84
Mahesh Chitrao	71	Paula S. Newman	243
Martin S. Chodorow	144	J. Terry Nutter	101
Kenneth W. Church	136	Duane Olawsky	109
Koenraad DeSmedt	77	Beatrice T. Oshika	203
Eva Ejerhed	219	Ping Peng	66
Olivier Emorine	115	Lisa F. Rau	129
Bruce Evans	203	Yael Ravin	144
Martha Evens	101	Elaine A. Rich	18, 186
Edward A. Fox	101	Stephen D. Richardson	195
Rudi Gebruers	168	Omneya A. Rizk	84
Ralph Grishman	66, 71	Dietmar Roesner	25
Philip J. Hayes	9	Rebecca Root	109
Lynette Hirschman	49	Karen L. Ryan	109
Jerry R. Hobbs	235	Howard E. Sachar	144
Sven Hurum	58	Sunil Shende	178
Paul S. Jacobs	129	Jonathan Slocum	228
Karen Jensen	152	Norman Sondheimer	176
Jugal Kalita	178	Janet Tom	203
Laura E. Knecht	9	Brigette van Berkel	77
Francois-Michel Lang	49	Giovanni B. Varile	160