29th Annual Meeting of the Association for Computational Linguistics

Proceedings of the Conference

18-21 June 1991
University of California
Berkeley, California, USA

Published by the Association for Computational Linguistics

© 1991, Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Donald E. Walker (ACL)
Bellcore, MRE 2A379
445 South Street, Box 1910
Morristown, NJ 07960-1910, USA

PREFACE

This volume contains the papers prepared for the *29th Annual Meeting of the Association for Computational Linguistics*, held 18–21 June 1991 in Berkeley, California. Also, for the first time the program included a number of presentations by students describing interesting dissertation work in progress. Their papers are gathered in a separate section at the end of this Proceedings.

The program committee found its job to be quite difficult this year because of the extraordinarily high quality of the submissions we received. Because of the high overall quality, we have expanded the number of papers we accepted this year, and even so we felt that some papers were rejected that would have met the standards of past years. It gives all of us great pleasure to see the field evolve and its standards of excellence increase. The entire computational linguistics research community deserves congratulations for this encouraging development. We feel this program demonstrates excellent progress in all areas of computational linguistics, and it reflects the increasingly international character of the research community, with contributions from Europe, Asia, the Middle East, and North America. We hope that you find these papers as interesting and exciting as we did.

As Program Chair, I wish thank our invited speakers, Sue Atkins, Charles Fillmore, and Jun-ichi Tsujii for their contributions to the program. I also thank Cecile Paris for organizing the tutorial sessions, and Jaime Carbonell, Julia Hirschberg, Nancy Ide, Kathleen McCoy, Johanna Moore, and Yorick Wilks for giving the tutorials. I also wish to extend my gratitude to Peter Norvig for the local arrangements, and to Sandra Newton for organizing the exhibit program, and of course, to Don Walker and Betty Walker for their tireless efforts on behalf of the *ACL*.

Finally, I wish to thank my fellow 1991 program committee members for their efforts in reading the 165 submitted papers: Ken Church, *AT&T Bell Labs and USC/ISI*; Robin Cohen, *University of Waterloo*; Erhard Hinrichs, *University of Tübingen*; Eduard Hovy, *USC/ISI*; Robert Ingria *BBN Systems and Technologies*; Yasuhiro Katagiri, *NTT Basic Research Laboratories*; Diane Litman, *Columbia University*; K. Vijay-Shanker, *University of Delaware*; Meg Withgott, *XEROX Palo Alto Research Center*; and Henk Zeevat, *University of Amsterdam.*

Douglas E. Appelt, *SRI International*
Chair, Program Committee

CONFERENCE PROGRAM

TUESDAY, 18 JUNE

9:00–12:30 **TUTORIAL SESSIONS**

Natural Language Generation
Kathleen McCoy and Johanna Moore

Intonation in Spoken Language Systems
Julia Hirschberg

2:00–5:30 **TUTORIAL SESSIONS**

Computational Linguistics Methodologies for Humanities Computing
Nancy M. Ide

Machine Translation: An In-Depth Tutorial
Jaime Carbonell and Yorick Wilks

WEDNESDAY, 19 JUNE

8:30–8:45 **OPENING REMARKS AND ANOUNCEMENTS**

8:45–9:10 *Resolution of Collective-Distributive Ambiguity Using Model-Based Reasoning*
Chinatsu Aone

9:10–9:35 *Inclusion, Disjointness and Choice: The Logic of Linguistic Classification*
Bob Carpenter & Carl Pollard

9:35–10:00 *Event-Building through Role-Filling and Anaphora Resolution*
Greg Whittemore, Melissa Macpherson & Greg Carlson

10:20–10:45 *Toward a Plan-Based Understanding Model for Mixed-Initiative Dialogues*
Hiroaki Kitano & Carol Van Ess-Dykema

10:45–11:10 *An Algorithm for Plan Recognition in Collaborative Discourse*
Karen E. Lochbaum

11:10–11:30 *A Three-Level Model for Plan Exploration*
Lance A. Ramshaw

11:30–11:50 *A Tripartite Plan-Based Model of Dialogue*
Lynn Lambert & Sandra Carberry

12:00–1:30 **STUDENT SESSION 1**

Logical Form of Complex Sentences in Task-Oriented Dialogues
Cecile T. Balkanski, Harvard University

Action Representation for NL Instructions
Barbara Di Eugenio, University of Pennsylvania

Extracting Semantic Roles from a Model of Eventualities
Sylvie Ratté, University of Québec at Montréal

Case Revisited: In the Shadow of Automatic Processing of Machine-Readable Dictionaries
Fuliang Weng, New Mexico State University

12:00–1:30 **STUDENT SESSION 2**

Discovering the Lexical Features of a Language
Eric Brill, University of Pennsylvania

Lexical Disambiguation: Information Sources and their Statistical Realization
Ido Dagan, Technion

Non-Literal Word Sense Identification through Semantic Network Path Schemata
Eric Iverson & Stephen Helmreich, New Mexico State University

1:40–2:05 *Discourse Relations and Defeasible Knowledge*
Alex Lascarides & Nicholas Asher

2:05–2:30 *Some Facts about Centers, Indexicals, and Demonstratives*
Rebecca J. Passonneau

2:30–2:55 *Type-Raising and Directionality in Combinatory Grammar*
Mark Steedman

3:15–3:40 *Efficient Incremental Processing with Categorial Grammar*
Mark Hepple

3:40–4:05 *Compose-Reduce Parsing*
Henry S. Thompson, Mike Dixon & John Lamping

4:05–4:30 *LR Recursive Transition Networks for Earley and Tomita Parsing*
Mark Perlin

4:50–5:15 *Polynomial Time and Space Shift-Reduce Parsing of Arbitrary Context-Free Grammars*
Yves Schabes

5:15–5:40 *Head Corner Parsing for Discontinuous Constituency*
Gertjan van Noord

5:40–6:05 *The Acquisition and Application of Context Sensitive Grammar for English*
Robert F. Simmons & Yeong-Ho Yu

THURSDAY, 20 JUNE

8:30–8:55 *Two Languages Are More Informative than One*
Ido Dagan, Alon Itai & Ulrike Schwall

8:55–9:20 *Learning Perceptually-Grounded Semantics in the L_0 Project*
Terry Regier

9:20–9:45 *Subject-Dependent Co-occurrence and Word Sense Disambiguation*
Joe A. Guthrie, Louise Guthrie, Yorick Wilks & Homa Aidinejad

9:45–10:10 *A System for Translating Locative Prepositions from English into French*
Nathalie Japkowicz & Janyce M. Wiebe

10:30–10:55 *Translation by Quasi Logical Form Transfer*
Hiyan Alshawi, David Carter, Manny Rayner & Björn Gambäck

10:55–12:00 **Linguistic Problems and Extra-Linguistic Problems in Machine Translation (INVITED TALK)**
Jun-ichi Tsujii, UMIST

12:00–1:30 **STUDENT SESSION 3**

Collaborating on Referring Expressions
Peter A. Heeman, University of Toronto

Conceptual Revision for Natural Language Generation
Ben E. Cline, Virginia Polytechnic Institute & State University

Modifying Beliefs in a Plan-Based Dialogue Model
Lynn Lambert, University of Delaware

Resolving a Pragmatics Prepositional Phrase Attachment Ambiguity
Christine H. Nakatani, University of Pennsylvania

12:00–1:30 **STUDENT SESSION 4**

Current Research in the Development of a Spoken Language System using PARSEC
Carla B. Zoltowski, Purdue University

Syntactic Graphs and Constraint Satisfaction
Jeff Martin, University of Maryland

An Incremental Connectionist Phrase Structure Parser
James Henderson, University of Pennsylvania

1:45–2:05 *Aligning Sentences in Parallel Corpora*
Peter F. Brown, Jennifer C. Lai & Robert L. Mercer

2:05–2:25 *A Program for Aligning Sentences in Bilingual Corpora*
William A. Gale & Kenneth W. Church

2:25–2:50 *Experiments and Prospects of Example-Based Machine Translation*
Eiichiro Sumita & Hitoshi Iida

3:10–3:35 *Resolving Translation Mismatches with Information Flow*
Megumi Kameyama, Ryo Ochitani & Stanley Peters

3:35–4:00 *Automatic Noun Classification by Using Japanese-English Word Pairs*
Naomi Inoue

4:20–4:45 *Automatic Acquisition of Subcategorization Frames from Untagged Text*
Michael R. Brent

4:45–5:10 *Multiple Default Inheritance in a Unification-Based Lexicon*
Graham Russell, John Carroll & Susan Warwick-Armstrong

5:10–5:35 *Metaphoric Generalization through Sort Coercion*
Ellen Hays & Samuel Bayer

5:35–6:00 *Structural Ambiguity and Lexical Relations*
Donald Hindle & Mats Rooth

FRIDAY, 21 JUNE

8:30–8:55 *Strategies for Adding Control Information to Declarative Grammars*
Hans Uszkoreit

8:55–9:20 *Finite-State Approximation of Phrase Structure Grammars*
Fernando C.N. Pereira & Rebecca N. Wright

9:20–9:45 *Feature Logic with Weak Subsumption Constraints*
Jochen Dörre

10:05–11:05 **Word Meaning: Starting where the MRDs Stop (INVITED TALK)**
Charles Fillmore and Sue Atkins

1:10–1:35 *Word Sense Disambiguation using Statistical Methods*
Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra & Robert L. Mercer

1:35–2:00 *A Stochastic Process for Word Frequency Distributions*
Harald Baayen

2:00–2:25 *From N-Grams to Collocations: An Evaluation of Xtract*
Frank A. Smadja

2:45–3:10 *Predicting Intonational Phrasing from Text*
Michelle Q. Wang & Julia Hirschberg

3:10–3:35 *A Preference-First Language Processor Integrating the Unification Grammar and Markov Language Model for Speech Recognition Applications*
Lee-Feng Chien, K.J. Chen & Lin-Shan Lee

3:35–4:00 *Factorization of Language Constraints in Speech Recognition*
Roberto Pieraccini & Chin-Hui Lee

4:20–4:45 *Constraint Projection: An Efficient Treatment of Disjunctive Feature Descriptions*
Mikio Nakano

4:45–5:05 *Quasi-Destructive Graph Unification*
Hideto Tomabechi

5:05–5:25 *Unification with Lazy Non-Redundant Copying*
Martin C. Emele

TUTORIALS

Natural Language Generation

Kathleen McCoy, *University of Delaware*, and Johanna Moore, *University of Pittsburgh*

The ability to generate natural language utterances is an important component of many intelligent systems (expert systems, intelligent tutoring systems, advice-giving systems). This tutorial will provide an in-depth survey of the branch of computational linguistics known as natural language generation. Most of the work in natural language processing has concentrated on understanding text. Instead, we look at the problems involved in generating text. Generation brings up issues not apparent in understanding. The task of an understander is to recognize which choice has been taken. In contrast, a generator must decide why to make one choice over another. Considering generation forces the researcher to come to terms with issues concerning what kind of information must be available to the generation component, where that information may be obtained, and how information should be presented to different users in different situations. In this tutorial we concentrate on a portion of the generation process known as *text planning*, which is responsible for deciding what is to be said and how it is to be structured. We look at how the content of a text can be chosen (including topics in user modeling and text planning formalisms) and how texts should be structured in a coherent fashion (including topics on text structure and coherence, pragmatics, focus of attention).

Intonation in Spoken Language Systems

Julia Hirschberg, *AT&T Bell Laboratories*

Current interest in spoken language systems has focused attention on potential interfaces between traditional concerns of Natural Language Processing with syntactic, semantic and discourse/pragmatics representation and analysis – and traditional concerns of speech scientists and engineers with speech recognition and synthesis. One area of mutual interest is intonational variation. How do intonational features such as phrasing and prominence interact with syntactic, semantic and discourse factors to shape the overall 'meaning' of an utterance? (Can parsers be designed to parse intonational features along with lexical items? Can intonation disambiguate among possible semantic interpretations of a sentence?) How can knowledge of intonational regularities improve speech recognition techniques and provide more natural-sounding synthetic speech? (Can intonational information be incorporated into recognition hypotheses? Can likely intonational features be reliably predicted from text to approximate human intonation in synthetic speech?) This tutorial will survey (a) current empirical and theoretical research on the contribution of intonation to utterance interpretation, (b) methods of prosodic analysis from speech corpora, (c) alternative approaches to intonational description and representation, and (d) current and potential applications to speech generation systems, text-to-speech systems and speech recognition. The tutorial will be extensively illustrated with examples from natural and synthetic speech.

Computational Linguistics Methodologies for Humanities Computing

Nancy M. Ide, *Vassar College*

Recently, panels and sessions at COLING, and conferences of the Association for Computational Linguistics, the Association for Computers and the Humanities, and the Association for Literary and Linguistic Computing have addressed the increasing merging of methodologies in the fields of computational linguistics and humanities computing. On the one hand, computational linguists are devoting considerable attention to statistical and other quantitative measures traditionally used in humanities computing. Also, work with large text corpora, long the central activity in humanities computing, is becoming an important area for computational linguistics. Computational linguists are now beginning to consider texts, and even literary texts, as an object of study and a rich source of information about the phenomena of language and discourse. On the other hand, humanists are turning to methods for morphological, syntactic, and semantic analysis developed by computational linguists to enhance their strategies for literary and linguistic studies. This tutorial will describe work which falls at the intersection of the fields of computational linguistics and humanities computing, either in methodology or use of materials, and show how these methods and materials benefit both disciplines. In particular, work in the areas of computational lexicology and lexicography, corpora and corpus linguistics, statistical models and methods for language and text analysis, and content analysis will be considered.

Machine Translation: An In-Depth Tutorial

Jaime Carbonell, *Carnegie Mellon University*, and Yorick Wilks, *New Mexico State University*

Machine Translation (MT) is the area of computational linguistics with the longest history and with the largest volume of dedicated R&D resources on the global scene. After reviewing the primary objectives and accomplishments of MT in its 40-year history, the major MT paradigms will be presented in some detail, including syntactic transfer, semantic transfer, and interlingua-based approaches. Then, we will discuss the appropriateness of these methods to different application areas, including technical vs nontechnical text, specialized domains vs general text, multilingual vs bilingual requirements, spontaneous discourse vs prepared text, and full-translation vs text scanning vs fact extraction. We will also touch upon evaluation of MT systems and recent developments in MT such the re-emergence of statistical approaches, making knowledge-based interlingual MT systems practical, and the integration of MT with other technologies such as document production, optical character recognition, and speech understanding.

TABLE OF CONTENTS

STUDENT SESSION TABLE OF CONTENTS

PREFACE TO THE STUDENT SESSION PAPERS

At the end of this volume are the papers accepted for short presentations at the student sessions of the *29th Annual Meeting of the Association for Computational Linguistics*. Unlike the main conference sessions, where the emphasis is on completed work, the student sessions emphasize work *in progress*, judged on the basis of creativity, insight, and promise.

These student sessions are an experiment that might, if successful, fulfill several desirable goals: (1) allowing student members to hone their presentation skills in front of a group not intimately familiar with their work; (2) providing student members with reactions to their research from the community at large, at a point where such reactions could make a significant difference; and (3) acting as "ice breakers," giving student members an opportunity to meet and talk with one another and with more senior researchers. If the student sessions have even partly fulfilled these goals, then our efforts have been worthwhile.

Our thanks go to all the students who submitted papers to the session and to the hard-working and responsive members of the Student Session Committee for their efforts in planning the sessions and in serving as reviewers: Dania Egedi, *Duke University*; Jong-Gyun Lim, *Columbia University*; Susan McRoy, *University of Toronto*; Jeff Siskind, *Massachusetts Institute of Technology*; David Traum, *University of Rochester*; and Barbara Vauthey, *New York University and Fribourg University*. We are also grateful to the following regular *ACL* members, who spent hours reviewing and commenting on the papers: Sandra Carberry, *University of Delaware*; Mark Liberman, *University of Pennsylvania*; Terry Nutter, *Virginia Polytechnic Institute*; Bill Rapaport, *SUNY Buffalo*; Tomek Strzalkowski, *New York University*; and Kent Wittenberg, *Bellcore and MCC*.

The student sessions would not have been possible without the initiative and support of the *ACL* Executive Committee and the invaluable assistance of Marti Hearst and Peter Norvig. We owe a special debt of thanks to Doug Appelt and Don Walker for their advice and encouragement.

Bonnie Lynn Webber, *University of Pennsylvania*
Chair, Student Session Program Committee

Philip Resnik, *University of Pennsylvania*
Student Session Coordinator

AUTHOR INDEX

STUDENT AUTHOR INDEX