

**37th
Annual Meeting
of the
Association for
Computational Linguistics**

Proceedings of the Conference

20-26 June 1999
University of Maryland
College Park, Maryland, USA

Published by the Association for Computational Linguistics
<http://www.aclweb.org/>

Production and Manufacturing by
Omipress, Inc.
Post Office Box 7214
Madison, WI 53707-7314

Copyright © 1999 by Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics

<http://www.aclweb.org>

or

Morgan Kaufmann Publishers

San Francisco

<http://www.mkp.com>

orders@mkp.com

1-800-745-7323

1-800-814-6418 (fax)

37TH Annual Meeting of the Association for Computational Linguistics
ISBN: 1-55860-609-2

PREFACE

This volume contains the papers accepted for presentation at the 37th Annual Meeting of the Association for Computational Linguistics, held at the University of Maryland from 20th to 26th June 1999.

As co-chairs, we had two particular objectives for the program this year. First, we wanted to acknowledge the increasing worldwide interest in computational linguistics, and mirror the move towards true internationalism that is underway within the ACL itself. Second, we wanted to reflect the broader connections that computational linguistics has with other research communities by increasing the scope of the material covered. You will see the results of our pursuit of these objectives throughout the program and its organisation: the 80 papers collected here for presentation were drawn from over 320 received, with the authors representing 29 countries and the reviewers drawn from 25 countries.

The larger-than-usual size of the conference this year is at least in part due to our incorporation of community-proposed theme sessions. In an iteration prior to the call for papers itself, we issued a call for theme session proposals. This was our attempt to break with the somewhat traditional set of paper topic categories that one tends to think of at ACL conferences. As a way of highlighting real areas of topical interest in the community, we think the experiment worked extremely well, despite the inevitable hiccups that accompany the introduction of any new idea. But you can be the judge of this: the ACL executive committee and next year's program chairs will be interested to hear what you think of this innovation.

Selecting the papers we could fit into the program from the high number of submissions received was not an easy task. In total over 200 reviewers were involved in the reviewing process, reporting through a thematic session committee with 12 members and a final program committee of six persons. Without the help of all these people, managing the large number of papers submitted would simply not have been feasible. All those who contributed are named on the following pages, but we would like in particular here to express our thanks to Stephen Abney (AT&T Labs – Research) and Graeme Ritchie (University of Edinburgh, UK), who helped in reviewing the theme proposals at extremely short notice, and the members of the program committee: Nicoletta Calzolari (Istituto di Linguistica Computazionale del CNR, Italy); Barbara Di Eugenio (University of Illinois at Chicago, USA); Hwee Tou Ng (DSO National Laboratories, Singapore); Cécile Paris (CSIRO Sydney, Australia); Keh-Yih Su (Behavior Design Corporation, Taiwan); and K. Vijay-Shanker (University of Delaware, USA). The program committee spent two intensive days at a meeting in New Jersey reaching the final decisions, enduring a snowstorm in the process.

There are, of course, a great many other people without whom this event would not be possible. We extend a special thanks to our invited speakers, Marti Hearst, Sadaoki Furui and George Miller, all of whom offer varying perspectives on computational linguistics and its links to other areas of endeavour. We are indebted to a number of individuals for taking care of specific parts of the conference program: Susan Armstrong (ISSCO, Switzerland) put together a rich workshops program; Melanie Baljko (University of Toronto, Canada) and Anna Korhonen (University of Cambridge, UK) chaired the student session committee; and Richard Sproat (AT&T Labs, USA) organised the excellent tutorial program. Maria Milosavljevic (CSIRO Sydney, Australia) built and maintained the main ACL99 web site, which served as our principal information channel.

On the ground at the University of Maryland, Bonnie Dorr managed the local arrangements, assisted by Cecilia Kullman, Edna Walker, Amy Weinberg, Gina Levow and Philip Resnik; David Traum arranged the demonstration sessions, and John White was responsible for sponsorship organisation. They have worked hard to provide you with what promises to be a very enjoyable conference experience.

Throughout the last year, we as co-chairs have constantly relied on the experience and help of both Kathy McCoy, ACL Secretary-Treasurer, and Priscilla Rasmussen, ACL Business Manager; the ACL Executive Committee provided useful advice especially in the early stages. At the end of the day, however, our greatest debt is to those who provided us with administrative support at our two host organisations: Joanne Avery and Cheryl Bock at AT&T and Debbie Whittington at Macquarie University's Language Technology Group. Their always speedy response to our last-minute requests for assistance made all the difference.

Welcome to ACL 99!

Robert Dale and Ken Church
Program Co-chairs
May 1999

PREFACE TO THE STUDENT SESSION PAPERS

The proceedings of ACL'99 wouldn't be complete without the extended abstracts of the students presenting their work in the special student sessions. Unlike the regular papers presented in the main session, students were encouraged to present work in progress that includes great ideas for future research.

We received 30 submissions from 12 countries. After thorough consideration, we accepted 10 of the submitted papers. Each submission was reviewed by at least two student reviewers and one non-student reviewer.

We want to thank the 31 members of the program committee for being precise and fair in their assessments and for making our lives easier by returning all reviews on time.

The student members of the ACL 1999 Student Session Program Committee were:

- Antonio Branco, Universidade de Lisboa, Portugal
- Hua Cheng, University of Edinburgh, Scotland, UK
- Jeanne Fromer, Massachusetts Institute of Technology, USA
- Marsal Gavalda, Carnegie Mellon University, USA
- Ulrich Germann, ISI, USA
- Daqing He, University of Edinburgh, Scotland, UK
- Sylvia Knight, University of Cambridge, UK
- Yuval Krymolowski, Bar-Ilan University, Israel
- Oi Yee Kwong, University of Cambridge, UK
- Dragomir Radev, Columbia University, USA
- James Shaw, Columbia University, USA
- Ivelin Stoianov, University of Groningen, The Netherlands
- Scott Thede, Purdue University, USA
- Peter Vanderheyden, University of Waterloo, Canada
- Jakub Zavrel, Tilburg University, The Netherlands
- Klaus Zechner, Carnegie Mellon University, USA

The following were the non-student members of the committee:

- Amit Bagga, Duke University, USA
- Roy Byrd, IBM, USA
- Lauri Carlson, University of Helsinki, Finland
- Christy Doran, University of Pennsylvania, USA
- Mark Dras, MRI, Macquarie University, Australia
- Kristiina Jokinen, ATL ITR, Japan
- Alistair Knott, University of Edinburgh, Scotland, UK
- Mark Lewellen, Georgetown University, USA
- Vera Lucia Strube de Lima, Pontificia Universidade Católica do Rio Grande do Sul, Brazil
- Chris Mellish, University of Edinburgh, Scotland, UK
- John Nerbonne, University of Groningen, The Netherlands
- Chris Reed, Brunel University, UK
- Paul Schmidt, Johannes Gutenberg-University, Germany
- Claudia Soria, University of Pisa, Italy
- Manfred Stede, Technische Universität Berlin, Germany

We want to thank last year's organizers: Maria Milosavljevic, CSIRO, Australia and Dragomir R. Radev, Columbia University, USA for their help and for providing us with a large volume of supporting materials that made it possible to handle the organization of this year's Student Session.

Finally, we want to thank the organizers of the main conference and the ACL Executive Committee, more specifically Robert Dale, Kenneth Church and Kathy McCoy.

Student Session Co-Chairs:

Melanie Baljko, University of Toronto, Canada
Anna Korhonen, University of Cambridge, UK

List of Reviewers

Program Committee:

Nicoletta Calzolari, Barbara Di Eugenio, Hwee Tou Ng, Cécile Paris, Keh-Yih Su, K. Vijay-Shanker.

Theme Sessions Committee:

James Allan, Jill Burstein, W. Bruce Croft, Laurence Danlos, Gregory Grefenstette, Jan Hajic, Tony Hartley, Julia Hirschberg, Lynette Hirschman, Nancy Ide, Douglas A. Jones, Alistair Knott, Alon Lavie, Claudia Leacock, Diane Litman, Masaaki Nagata, Douglas Oard, Boyan Onyshkevych, David Palmer, Cécile Paris, Owen Rambow, Philip Resnik, Carolyn Penstein Rose, Hinrich Schütze, Max Silberztein, Elke Teich, Marilyn Walker, Bonnie Webber, Sandra Williams.

Reviewers:

Anne Abeille (Université de Paris 7), Steven Abney (AT&T Labs – Research), Jan Alexandersson (DFKI), James Allan (University of Massachusetts, Amherst), Nicholas Asher (University of Texas).

Lisa Ballesteros (University of Massachusetts, Amherst), Srinivas Bangalore (AT&T Labs – Research), Tilman Becker (DFKI), Nuria Bel (University of Barcelona), Javier Blanco (Autonomous University of Barcelona), Branimir Boguraev (IBM T. J. Watson Research Center), Gosse Bouma (Rijksuniversiteit Groningen), Michael Brent (Johns Hopkins University), Chris Brew (University of Edinburgh), Ted Briscoe (University of Cambridge), Rebecca Bruce (University of North Carolina at Asheville), Federica Busa (Brandeis University).

Jamie Callan (University of Massachusetts, Amherst), Nick Campbell (ATR Interpreting Telecommunications Research Lab), Sandee Carberry (University of Delaware), Jaime Carbonell (Carnegie Mellon University), Claire Cardie (Cornell University), Jean Carletta (University of Edinburgh), Lauri Carlson (Helsinki University), Rolf Carlson (KTH), Lynn Carlson, Bob Carpenter (Lucent Technologies), John Carroll (University of Sussex), David Carter (SRI Cambridge), Jason S. Chang (National Tsing-Hua University), Jing-Shin Chang (Behavior Design Corporation), Eugene Charniak (Brown University), Ciprian Chelba (Johns Hopkins University), Hsin-Hsi Chen (National Taiwan University), Stanley Chen (CMU), Keh-Jiann Chen (Academia Sinica), Tung-Hui Chiang (CCL/ITRI), Key-Sun Choi (KAIST), Michael Collins (AT&T Labs – Research), Mark Core (University of Rochester), Stephen Crain (University of Maryland), Dan Cristea (University A. I. Cuza Iasi), Krzysztof Czuba (Carnegie Mellon University).

Walter Daelemans (Tilburg University), Nils Dählback (Linköping University), Morena Danieli (CSELT (CF/VR)), Satya Dharanipragada (IBM Watson), Anne Dister (University of Liege), John Dowding (SRI).

Jeff Elman (UCSD), Cedrick Fairon (University of Marne-la-Vallée), Christiane Fellbaum (Princeton University), Charles J. Fillmore (International Computer Science Institute), Dan Flickinger (Stanford University), Robert Frank (Johns Hopkins University), Norbert Fuchs (University of Zurich), Pascale Fung (HKUST).

Marsal Gavaldà (Carnegie Mellon University), Laurie Gerber (SYSTRAN Software Inc.), Ted Gibson (MIT), Michael Glass (IIT), Andrew Golding (MERL), Nancy Green (Carnegie Mellon University), Greg Grefenstette (Xerox Research Centre Europe), Maurice Gross (Université de Paris 7), Jin Guo (Motorola).

Udo Hahn (Universitaet Freiburg), Masahiko Haruno (ATR), Kōiti Hasida (ETL), Marti Hearst (UC Berkeley), Peter Heeman (Oregon Graduate Institute), James Henderson (University of Exeter), Donald Hindle (AT&T Labs – Research), Barbora Hladka (Charles University), Eduard Hovy (Information Sciences Institute), Changning Huang (Tsinghua University).

Hitoshi Iida (Sony), Pierre Isabelle (Xerox Research Center Europe), Masato Ishizaki (Japan Advanced Institute of Science and Technology).

Christian Jacquemin (LIMSI), Ian Johnson (Sharp Laboratories of Europe), Mark Johnson (Brown University), Kristiina Jokinen (ATR), Douglas A. Jones (Dept. of Defense), Aravind Joshi (University of Pennsylvania), Dan Jurafsky (University of Colorado Boulder).

Noriko Kando (NACSIS), Ronald Kaplan (Xerox Palo Alto Research Center), Jussi Karlgren (Helsinki University and SICS), Michael Kearns (AT&T Labs – Research), Andrew Kehler (SRI International), Richard Kittredge (Montreal University), Kevin Knight (USC/ISI), Hanae Koiso (National Language Research Institute), Steven Krauwer (Utrecht University), Sadao Kurohashi (Kyoto University), Natasha Kurtonina (University of Pennsylvania).

Lillian Lee (Cornell University), Geunbae Lee (Pohang University of Science and Technology), Mun Kew Leong (Kent Ridge Digital Labs), Leonardo Lesmo (Universita' di Torino), Lori Levin (Carnegie Mellon University), David Lewis (AT&T Labs – Research), Hang Li (NEC Corporation), Dekang Lin (University of Manitoba), Susann LuperFoy (Information Extraction and Transport).

Christopher Manning (University of Sydney), Alec Marantz (Massachusetts Institute of Technology), Daniel Marcu (Information Sciences Institute/USC), Paul Martin (Sun Microsystems Laboratories), Yuji Matsumoto (NAIST), John Maxwell (Xerox Palo Alto Research Center), David McGee (Oregon Graduate Institute), Dan Melamed (West Group), Paola Merlo (University of Geneva), David Miller (BBN Technologies), Mehryar Mohri (AT&T Labs – Research), Johanna Moore (University of Edinburgh), Sung Hyon Myaeng (Chungnam National University).

Jian-Yun Nie (University of Montreal), Anton Nijholt (University of Twente), Sergei Nirenburg, Eric Nyberg (Carnegie Mellon University).

Kemal Oflazer (Bilkent University/NMSU-CRL), Antoine Ogonowski (LexiQuest), Karel Oliva (University of the Saarland).

Martha Palmer (University of Pennsylvania), Patrick Paroubek (LIMSI), Jan Pedersen (Infoseek Corporation), Carol Peters (IEI-CNR Pisa), Vito Pirrelli (ILC-CNR Pisa), Massimo Poesio (University of Edinburgh), Scott Prevost (FX Palo Alto Laboratory), Gábor Prószéky (Morphologic, Inc.), Stephen Pulman (SRI and Cambridge University).

Elisabete Ranchhod (University of Lisboa), Victor Raskin (Purdue University), Adwait Ratnaparkhi (IBM), Manny Rayner (SRI International), Terry Regier (University of Chicago), Norbert Reithinger (DFKI GmbH Saarbruecken), Antoinette Renouf (University of Liverpool), Ellen Riloff (University of Utah), James Rogers (University of Central Florida), Laurent Romary (Loria/CNRS), Mats Rooth (University of Stuttgart), Carolyn Rose (University of Pittsburgh), Alexandr Rosen (Charles University), Dan Roth (University of Illinois at Urbana/Champaign), John Rotondo (AT&T Labs – Research).

David Sadek (CNET/France Telecom), Mark Sanderson (University of Sheffield), Giorgio Satta (Universita' di Padova), Michael Schultz (University of Pennsylvania), Stephanie Seneff (MIT), Jean Senellart (Université de Paris 7), Candace Sidner (Lotus Development Corporation), Amit Singhal (AT&T Labs – Research), Alan Smeaton (Dublin City University), Ronnie Smith (East Carolina University), Harold Somers (UMIST, Manchester), Virach Sornlertlamvanich (NECTEC), Michael Spivey (Cornell University), Richard W. Sproat (AT&T Labs – Research), Bangalore Srinivas (AT&T Labs – Research), Ed Stabler (UCLA), Manfred Stede (Technical University of Berlin), Mark Steedman (University of Edinburgh), Gees Stein, Suzanne Stevenson (Rutgers University), Matthew Stone (Rutgers University), Michael Strube (University of Pennsylvania), Tomek Strzalkowski (General Electric), Mao-Song Sun (Tsinghua University).

Hozumi Tanaka (Tokyo Institute of Technology), Jacques Terken (IPO Eindhoven University of Technology), Jun'ichi Tsujii (University of Tokyo).

Juan Uriagereka, Enric Vallduvi (Universitat Pompeu Fabrá), Keith van Rijsbergen (Glasgow University), Ellen Voorhees (NIST), Atro Voutilainen (University of Helsinki).

Bonnie Webber (University of Edinburgh), Amy Weinberg (University of Maryland), David Weir (University of Sussex), Janyce Wiebe (New Mexico State University), Mats Wiren (Telia Research), Richard Wojcik (Boeing), Karsten Worm (University of the Saarland), Dekai Wu (HKUST).

Mikio Yamamoto (University of Tsukuba), David Yarowsky, Daniel Zeman (Charles University), Teresa Zollo (University of Rochester).

Day 1 Wednesday 23rd June

8:45 – 9:00	Opening Remarks	
9:00 – 10:00	Invited Speaker: Marti Hearst: "Untangling Text Data Mining"	
10:00 – 10:25	WedA1: Distributional Similarity and Clustering	WedB1: Semantics and Pragmatics
	Lillian Lee Measures of Distributional Similarity	Bonnie Webber, Alistair Knott, Matthew Stone and Aravind Joshi Discourse Relations: A Structural and Presuppositional Account using Lexicalised TAG
10:25 – 10:50	Lillian Lee and Fernando Pereira Distributional Similarity Models: Clustering vs. Nearest Neighbors	Claire Gardent Unifying Parallels
10:50 – 11:25	Break	
11:25 – 11:50	WedA2: Inductive Methods	WedB2: Grammars
	Matthew Berland and Eugene Charuak Finding Parts in Very Large Corpora	Mark Dras A Meta-level Grammar: Redefining Synchronous TAG for Translation and Paraphrase
11:50 – 12:15	Eric Brill and Grace Ngai Man vs. Machine: A Case Study in Base Noun Phrase Learning	William Schuler Preserving Semantic Dependencies in Synchronous Tree Adjoining Grammar
12:15 – 12:40	Rebecca Hwa Supervised Grammar Induction using Training Data with Limited Constituent Information	Shuly Wintner Compositional Semantics for Linguistic Formalisms
12:40 – 14:20	Lunch	
14:20 – 14:45	WedA3: Corpus-based Acquisition of Semantic Information	WedB3: Natural Language Generation
	Mats Rooth, Stefan Riezler, Dettlef Prescher, Glenn Carroll and Franz Beil Inducing a Semantically Annotated Lexicon via EM-based Clustering	John A. Bateman Using Aggregation for Selecting Content when Generating Referring Expressions
14:45 – 15:10	Eric V. Siegel Corpus-based Linguistic Indicators for Aspectual Classification	James Shaw and Vasileios Hatzivassiloglou Ordering Among Premodifiers
15:10 – 15:35	Sharon A. Caraballo Automatic Construction of a Hypernym-labeled Noun Hierarchy from Text	Yael Dahan Netzer and Michael Elhadad Bilingual Hebrew-English Generation of Possessives and Partitives: Raising the Input Abstraction Level
15:35 – 16:10	Break	
16:10 – 16:35	WedA4: Lexical Disambiguation	WedB4: Statistical Language Processing
	Rada Mihalcea and Dan I. Moldovan A Method for Word Sense Disambiguation of Unrestricted Text	Radu Florian and David Yarowsky Dynamic Nonlocal Language Modeling via Hierarchical Topic-based Adaptation
16:35 – 17:00	Andrei Mikheev A Knowledge-free Method for Capitalized Word Disambiguation	Scott M. Thede and Mary P. Harper A Second-Order Hidden Markov Model for Part-of-Speech Tagging

Day 2 Thursday 24th June

8:45 – 9:10	ThuA1: Theme Session: Dialogue Management in Interactive Spoken Dialog Systems. Chairs: Diane Litman and Marilyn Walker	ThuB1: Theme Session: NLP Techniques for Cross-Language Information Retrieval. Chair: Doug Oard
9:10 – 9:35	A Stent, J Dowding, J M Gawron, E Owen Bratt and R Moore The CommandTalk Spoken Dialogue System	J. Scott McCarley Should we Translate the Documents or the Queries in Cross-Language Information Retrieval?
9:35 – 10:00	Alicia Abella and Allen L. Gorin Construct Algebra: Analytical Dialog Management	Hsin-Hsi Chen, Guo-Wei Bian and Wen-Cheng Lin Resolving Translation Ambiguity and Target Polysemy in Cross-Language Information Retrieval
10:00 – 10:35	M. Nakano, N. Miyazaki, J. Hirasawa, K. Dohsaka, and T. Kawabata Understanding Unsegmented User Utterances in Real-Time Spoken Dialogue Systems	Myung-Gil Jang, Sung Hyon Myaeng and Se Young Park Using Mutual Information to Resolve Query Translation Ambiguities and Query Term Weighting
	Break	
10:35 – 11:00	ThuA2: Theme Session: Discourse Tagging: Uses, Results and Applications. Chairs: M Walker, J Hirschberg and O Rambow	ThuB2: Theme Session: Parsing of inflective, agglutinative and/or free word order languages. Chair: Jan Hajic
11:00 – 11:25	Won Seug Choi, Jeong-Mi Cho and Jungyun Seo Analysis System of Speech Acts and Discourse Structures Using Maximum Entropy Model	Kemal Oflazer Dependency Parsing with an Extended Finite State Approach
11:25 – 11:50	Sherri L. Condon, Claude G. Cech and William R. Edwards Measuring Conformity to Discourse Routines in Decision-Making Interactions	Gábor Prószéky and Balázs Kis A Unification-based Approach to Morpho-syntactic Parsing of Agglutinative and Other (Highly) Inflectional Languages
11:50 – 13:45	Janyce M. Wiebe, Rebecca F. Bruce and Thomas P. O'Hara Development and Use of a Gold Standard Data Set for Subjectivity Classifications	Franz Beil, Glenn Carroll, Detlef Prescher, Stefan Riezler and Mats Rooth Inside-Outside Estimation of a Lexicalized PCFG for German
13:45 – 14:45	Lunch	
14:45 – 15:10	ThuA3: Statistical Language Processing Masaaki Nagata Invited Speaker: Sadaoaki Furui: "Automatic Speech Recognition and its Application to Information Extraction"	ThuB3: Semantics and Quantifier Scoping Alistair Willis and Suresh Manandhar Two Accounts of Scope Availability and Semantic Underspecification
15:10 – 15:35	Antal van den Bosch and Walter Daelemans Memory-based Morphological Analysis	Mark Steedman Alternating Quantifier Scope in CCG
15:35 – 16:10	Break	
16:10 – 16:35	ThuA4: Student Session: Spoken Language Kavita Thomas Designing a Task-based Evaluation Methodology for a Spoken Machine Translation System.	ThuB4: Student Session: Grammar, Parsing and Syntax Yusuke Miyao Packing of Feature Structures for Efficient Unification of Disjunctive Feature Structures
16:35 – 17:00	Edward C. Kaiser Robust, Finite-State Parsing for Spoken Language Understanding	Alexandra Kinyon Parsing Preferences with Lexicalized Tree Adjoining Grammars: Exploiting the Derivation Tree

Day 3 Friday 25th June

					FriB1: Information Retrieval
8:45 – 9:10	Diane J. Litman, Marilyn A. Walker and Michael S. Kearns Dekang Lin	Automatic Detection of Poor Speech Recognition at the Dialogue Level Automatic Identification of Non-compositional Phrases		Pascale Fung, Liu Xiaohu and Cheung Chi Shun Christian Jacquemin	Mixed Language Query Disambiguation Syntagmatic and Paradigmatic Representations of Term Variation
9:35 – 10:00	Lynette Hirschman, Marc Light, Eric Breck and John D. Burger	Deep Read: A Reading Comprehension System		Simon H. Corston-Oliver and William B. Dolan	Less is More: Eliminating Index Terms from Subordinate Clauses
10:00 – 10:35					Break
10:35 – 11:00	Stefan Kaufmann	Semantics, Discourse and Pragmatics Cohesion and Collocation: Using Context Vectors in Text Segmentation		Adrian Corduneanu	Statistical Language Processing A Pylomic Decision-Tree Language Model with Optimal Question Selection
11:00 – 11:25	Horacio Saggion	Using Linguistic Knowledge in Automatic Abstracting		Miriam Goldberg	An Unsupervised Model for Statistically Determining Coordinate Phrase Attachment
11:25 – 11:50	Joel R. Tetreault	Analysis of Syntax-based Pronoun Resolution Methods		Freddy Y. Y. Choi	A Flexible Distributed Architecture for NLP System Development and Use
11:50 – 13:20					Lunch
13:20 – 14:35					ACL Business Meeting
14:35 – 15:10					Break
					FriB3: Theme Session: Robust and Shallow Parsing. Chairs: Gregory Grefenstette, Carolyn Fenstein Rose and Alon Lavie
15:10 – 15:45	Jeffrey C. Reynar	Statistical Models for Topic Segmentation		Emmanuel Morin and Christian Jacquemin	Projecting Corpus-based Semantic Links on a Thesaurus
15:45 – 16:10	Daniel Marcu	A Decision-based Approach to Rhetorical Parsing		Maria Lapata	Acquiring Lexical Generalizations from Corpora: A Case Study for Diathesis Alternations
16:10 – 16:35	David L. Bean and Ellen Riloff	Corpus-based Identification of Non-Anaphoric Noun Phrases		Walter Kasper, Bernd Kiefer, Hans-Ulrich Krieger, C. J. Rupp and Karsten L. Worm	Charting the Depths of Robust Speech Parsing
16:35 – 17:00	Hideki Tanaka and Akio Yokoo	An Efficient Statistical Speech Act Type Tagging System for Speech Translation Systems		Mark G. Core and Lenhart K. Schubert	A Syntactic Framework for Speech Repairs and Other Disruptions

Day 4 Saturday 26th June

	SatA1: Theme Session: Computational Psycholinguistics. Chair: Philip Resnik		SatB1: Applications
8:45 – 9:10	Brian Roark and Mark Johnson	Efficient Probabilistic Top-down and Left-corner Parsing	Hwee Tou Ng, Chung Yong Lim and Jessica Li
9:10 – 9:35	Charles D. Yang	A Selectionist Theory of Language Acquisition	Teng Koo
9:35 – 10:00	Marielle Lange and Alain Content	The Grapho-phonological System of Written French: Statistical Analysis and Empirical Validation	Sergey V. Pakhomov
10:00 – 10:35	Break		
10:35 – 11:00	Jason Eisner and Giorgio Satta	Efficient Parsing for Bilexical Context-Free Grammars and Head Automaton Grammars	Sadao Kurohashi and Yasuyuki Sakai
11:00 – 11:25	Mark Hepple	An Early-style Predictive Chart Parsing Method for Lambek Grammars	Hitoshi Isahara and Kyoko Kanzaki
11:25 – 11:50	Bernd Kiefer, Hans-Ulrich Krieger, John Carroll and Rob Malouf	A Bag of Useful Techniques for Efficient and Robust Parsing	Patrick Caudal
11:50 – 13:45	Lunch		
13:45 – 14:10	Michael Collins, Jan Hajic, Lance Ramshaw and Christoph Tillmann	A Statistical Parser for Czech	SatB3: Bilingual Alignment
14:10 – 14:35	Don Blaheta and Eugene Charniak	Automatic Compensation for Parser Figure-of-Merit Flaws	Reinhard Rapp
14:35 – 15:10	Break		
15:10 – 15:35	Mark Johnson, Stuart Geman, Stephen Canon, Zhiyi Chi and Stefan Riezler	Estimators for Stochastic "Unification-based" Grammars	Philip Resnik
15:35 – 16:00	Steven Abney, David McAllester and Fernando Pereira	Relating Probabilistic Grammars and Automata	Automatic Identification of Word Translations from Unrelated English and German Corpora
16:00 – 17:00	Invited Speaker: George Miller: "The Lexical Component of Natural Language Processing"		

ACL 99 TUTORIALS SCHEDULE

SUNDAY, JUNE 20, 1999

Morning Tutorials (8:30 - 12:00)

- Computational Approaches to Gesture and Natural Language
Justine Cassell (MIT Media Lab)
- Lexicography for Computationalists
Adam Kilgarriff (University of Brighton)
Michael Rundell (Lexicography Master Class & Managing Editor, LDOCE)
- Symbolic Machine Learning for Natural Language Processing
Raymond Mooney (UT Austin)
Claire Cardie (Cornell University)

Afternoon Tutorials (1:30 - 5:00)

- Models of Memory and Analogy for Natural Language Learning and Processing
Antal van den Bosch (Tilburg University)
- Spoken Dialogue Systems
Bob Carpenter
Jennifer Chu-Carroll (both of Lucent Technologies, Bell Laboratories)
- Probabilistic Models for Artificial Intelligence
Michael Kearns (AT&T Labs)

ACL 99 Tutorials Chair:
Richard Sproat (AT&T Labs)

TABLE OF CONTENTS

Invited Talk: <i>Untangling Text Data Mining</i> Marti A. Hearst	3
Invited Talk: <i>Automatic Speech Recognition and its Application to Information Extraction</i> Sadaoki Furui	11
Invited Talk: <i>The Lexical Component of Natural Language Processing</i> George A. Miller.....	21
<i>Measures of Distributional Similarity</i> Lillian Lee.....	25
<i>Distributional Similarity Models: Clustering vs. Nearest Neighbors</i> Lillian Lee and Fernando Pereira.....	33
<i>Discourse Relations: A Structural and Presuppositional Account using Lexicalised TAG</i> Bonnie Webber, Alistair Knott, Matthew Stone and Aravind Joshi	41
<i>Unifying Parallels</i> Claire Gardent	49
<i>Finding Parts in Very Large Corpora</i> Matthew Berland and Eugene Charniak	57
<i>Man vs. Machine: A Case Study in Base Noun Phrase Learning</i> Eric Brill and Grace Ngai	65
<i>Supervised Grammar Induction using Training Data with Limited Constituent Information</i> Rebecca Hwa	73
<i>A Meta-level Grammar: Redefining Synchronous TAG for Translation and Paraphrase</i> Mark Dras	80
<i>Preserving Semantic Dependencies in Synchronous Tree Adjoining Grammar</i> William Schuler	88
<i>Compositional Semantics for Linguistic Formalisms</i> Shuly Wintner	96
<i>Inducing a Semantically Annotated Lexicon via EM-based Clustering</i> Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll and Franz Beil	104
<i>Corpus-based Linguistic Indicators for Aspectual Classification</i> Eric V. Siegel.....	112
<i>Automatic Construction of a Hypernym-labeled Noun Hierarchy from Text</i> Sharon A. Caraballo	120
<i>Using Aggregation for Selecting Content when Generating Referring Expressions</i> John A. Bateman.....	127

<i>Ordering Among Premodifiers</i> James Shaw and Vasileios Hatzivassiloglou	135
<i>Bilingual Hebrew-English Generation of Possessives and Partitives: Raising the Input Abstraction Level</i> Yael Dahan Netzer and Michael Elhadad	144
<i>A Method for Word Sense Disambiguation of Unrestricted Text</i> Rada Mihalcea and Dan I. Moldovan	152
<i>A Knowledge-free Method for Capitalized Word Disambiguation</i> Andrei Mikheev	159
<i>Dynamic Nonlocal Language Modeling via Hierarchical Topic-based Adaptation</i> Radu Florian and David Yarowsky	167
<i>A Second-Order Hidden Markov Model for Part-of-Speech Tagging</i> Scott M. Thede and Mary P. Harper	175
<i>The CommandTalk Spoken Dialogue System</i> Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt and Robert Moore	183
<i>Construct Algebra: Analytical Dialog Management</i> Alicia Abella and Allen L. Gorin	191
<i>Understanding Unsegmented User Utterances in Real-Time Spoken Dialogue Systems</i> Mikio Nakano, Noboru Miyazaki, Jun-ichi Hirasawa, Kohji Dohsaka, and Takeshi Kawabata	200
<i>Should We Translate the Documents or the Queries in Cross-Language Information Retrieval?</i> J. Scott McCarley	208
<i>Resolving Translation Ambiguity and Target Polysemy in Cross-Language Information Retrieval</i> Hsin-Hsi Chen, Guo-Wei Bian and Wen-Cheng Lin	215
<i>Using Mutual Information to Resolve Query Translation Ambiguities and Query Term Weighting</i> Myung-Gil Jang, Sung Hyon Myaeng and Se Young Park	223
<i>Analysis System of Speech Acts and Discourse Structures using Maximum Entropy Model</i> Won Seug Choi, Jeong-Mi Cho and Jungyun Seo	230
<i>Measuring Conformity to Discourse Routines in Decision-making Interactions</i> Sherri L. Condon, Claude G. Cech and William R. Edwards	238
<i>Development and Use of a Gold-Standard Data Set for Subjectivity Classifications</i> Janyce M. Wiebe, Rebecca F. Bruce and Thomas P. O'Hara	246
<i>Dependency Parsing with an Extended Finite State Approach</i> Kemal Oflazer	254
<i>A Unification-based Approach to Morpho-syntactic Parsing of Agglutinative and Other (Highly) Inflectional Languages</i> Gábor Prózszéky and Balázs Kis	261
<i>Inside-Outside Estimation of a Lexicalized PCFG for German</i> Franz Beil, Glenn Carroll, Detlef Prescher, Stefan Riezler and Mats Rooth	269

<i>A Part of Speech Estimation Method for Japanese Unknown Words using a Statistical Model of Morphology and Context</i> Masaaki Nagata.....	277
<i>Memory-based Morphological Analysis</i> Antal van den Bosch and Walter Daelemans.....	285
<i>Two Accounts of Scope Availability and Semantic Underspecification</i> Alistair Willis and Suresh Manandhar.....	293
<i>Alternating Quantifier Scope in CCG</i> Mark Steedman	301
<i>Automatic Detection of Poor Speech Recognition at the Dialogue Level</i> Diane J. Litman, Marilyn A. Walker and Michael S. Kearns.....	309
<i>Automatic Identification of Non-compositional Phrases</i> Dekang Lin.....	317
<i>Deep Read: A Reading Comprehension System</i> Lynette Hirschman, Marc Light, Eric Breck and John D. Burger	325
<i>Mixed Language Query Disambiguation</i> Pascale Fung, Liu Xiaohu and Cheung Chi Shun.....	333
<i>Syntagmatic and Paradigmatic Representations of Term Variation</i> Christian Jacquemin	341
<i>Less is More: Eliminating Index Terms from Subordinate Clauses</i> Simon H. Corston-Oliver and William B. Dolan	349
<i>Statistical Models for Topic Segmentation</i> Jeffrey C. Reynar	357
<i>A Decision-based Approach to Rhetorical Parsing</i> Daniel Marcu	365
<i>Corpus-based Identification of Non-Anaphoric Noun Phrases</i> David L. Bean and Ellen Riloff.....	373
<i>An Efficient Statistical Speech Act Type Tagging System for Speech Translation Systems</i> Hideki Tanaka and Akio Yokoo	381
<i>Projecting Corpus-based Semantic Links on a Thesaurus</i> Emmanuel Morin and Christian Jacquemin.....	389
<i>Acquiring Lexical Generalizations from Corpora: A Case Study for Diathesis Alternations</i> Maria Lapata.....	397
<i>Charting the Depths of Robust Speech Parsing</i> Walter Kasper, Bernd Kiefer, Hans-Ulrich Krieger, C. J. Rupp and Karsten L. Worm.....	405
<i>A Syntactic Framework for Speech Repairs and Other Disruptions</i> Mark G. Core and Lenhart K. Schubert	413

<i>Efficient Probabilistic Top-down and Left-corner Parsing</i> Brian Roark and Mark Johnson.....	421
<i>A Selectionist Theory of Language Acquisition</i> Charles D. Yang.....	429
<i>The Grapho-phonological System of Written French: Statistical Analysis and Empirical Validation</i> Marielle Lange and Alain Content	436
<i>Learning to Recognize Tables in Free Text</i> Hwee Tou Ng, Chung Yong Lim and Jessica Li Teng Koo	443
<i>A Semantically-derived Subset of English for Hardware Verification</i> Alexander Holt and Ewan Klein.....	451
<i>Efficient Parsing for Bilexical Context-Free Grammars and Head Automaton Grammars</i> Jason Eisner and Giorgio Satta.....	457
<i>An Earley-style Predictive Chart Parsing Method for Lambek Grammars</i> Mark Hepple.....	465
<i>A Bag of Useful Techniques for Efficient and Robust Parsing</i> Bernd Kiefer, Hans-Ulrich Krieger, John Carroll and Rob Malouf	473
<i>Semantic Analysis of Japanese Noun Phrases: A New Approach to Dictionary-based Understanding</i> Sadao Kurohashi and Yasuyuki Sakai	481
<i>Lexical Semantics to Disambiguate Polysemous Phenomena of Japanese Adnominal Constituents</i> Hitoshi Isahara and Kyoko Kanzaki	489
<i>Computational Lexical Semantics, Incrementality, and the So-called Punctuality of Events</i> Patrick Caudal.....	497
<i>A Statistical Parser for Czech</i> Michael Collins, Jan Hajic, Lance Ramshaw and Christoph Tillmann.....	505
<i>Automatic Compensation for Parser Figure-of-Merit Flaws</i> Don Blaheta and Eugene Charniak	513
<i>Automatic Identification of Word Translations from Unrelated English and German Corpora</i> Reinhard Rapp	519
<i>Mining the Web for Bilingual Text</i> Philip Resnik.....	527
<i>Estimators for Stochastic "Unification-based" Grammars</i> Mark Johnson, Stuart Geman, Stephen Canon, Zhiyi Chi and Stefan Riezler.....	535
<i>Relating Probabilistic Grammars and Automata</i> Steven Abney, David McAllester and Fernando Pereira.....	542
<i>Information Fusion in the Context of Multi-Document Summarization</i> Regina Barzilay, Kathleen R. McKeown and Michael Elhadad.....	550

<i>Improving Summaries by Revising Them</i> Inderjeet Mani, Barbara Gates, and Eric Bloedorn	558
Student Session Paper: <i>Designing a Task-based Evaluation Methodology for a Spoken Machine Translation System</i> Kavita Thomas.....	569
Student Session Paper: <i>Robust, Finite-State Parsing for Spoken Language Understanding</i> Edward C. Kaiser	573
Student Session Paper: <i>Packing of Feature Structures for Efficient Unification of Disjunctive Feature Structures</i> Yusuke Miyao.....	579
Student Session Paper: <i>Parsing Preferences with Lexicalized Tree Adjoining Grammars: Exploiting the Derivation Tree</i> Alexandra Kinyon.....	585
Student Session Paper: <i>Cohesion and Collocation: Using Context Vectors in Text Segmentation</i> Stefan Kaufmann.....	591
Student Session Paper: <i>Using Linguistic Knowledge in Automatic Abstracting</i> Horacio Saggion.....	596
Student Session Paper: <i>Analysis of Syntax-based Pronoun Resolution Methods</i> Joel R. Tetreault	602
Student Session Paper: <i>A Pylonic Decision-Tree Language Model with Optimal Question Selection</i> Adrian Corduneanu	606
Student Session Paper: <i>An Unsupervised Model for Statistically Determining Coordinate Phrase Attachment</i> Miriam Goldberg	610
Student Session Paper: <i>A Flexible Distributed Architecture for NLP System Development and Use</i> Freddy Y. Y. Choi.....	615
Student Session Paper: <i>Modeling Filled Pauses in Medical Dictations</i> Sergey V. Pakhomov	619