35th Annual Meeting of the Association for Computational Linguistics

and

8th Conference of the European Chapter of the Association for Computational Linguistics

Proceedings of the Conference

7-12 July 1997
Universidad Nacional de Educación a Distancia (UNED)
Madrid, Spain

Published by the Association for Computational Linguistics

The conference is sponsored by
Directorate General XIII (European Commission)
Universidad Nacional de Educación a Distancia
Universidad Autónoma de Madrid
Universidad Complutense de Madrid
Universidad Politécnica de Madrid

©1997, Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Morgan Kaufmann Publishers
340 Pine Street, 6th Floor
San Francisco, CA 94104, USA
E-Mail: orders@mkp.com
WWW: http://www.mkp.com

PREFACE

We are pleased to present you with this volume containing the papers accepted for presentation at the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics, July 7-12, 1997 at the Universidad Nacional de Educación a Distancia (UNED) Madrid, Spain.

In this, the first joint meeting of the American and European chapters of the Association for Computational Linguistics, you will find a wealth of high-quality papers, tutorials, and workshops on a wide range of topics related to natural language processing. This year's response to the Call for Papers was overwhelming - a 47% increase over the number of submissions to ACL'96. This dramatic increase in papers submitted indicates the field is thriving.

This volume contains 63 papers, which were chosen from the 264 that were submitted to the conference. With an acceptance rate of 24%, ACL-EACL'97 is one of the most selective international conferences of 1997.

This year, reviewing was conducted by a hierarchically structured program committee, consisting of five area chairs and 53 regular program committee members, supported by the invaluable assistance of 54 additional specialist reviewers. Their names are listed on the following pages. Reviewing was conducted blind, to all reviewers and area chairs.

All papers that remained controversial for some reason - or for which reviewers indicated low confidence in their judgements - were explicitly discussed at the meeting of the PC Chairs and the Area Chairs on March 2, 1997 in Brighton.

32 accepted papers come from North America (51%), 25 from Europe (40%) and 6 from Asia (9%). ACL-EACL 97 is a truely international conference: authors from 16 different countries contribute accepted papers. The top three contributing countries are: USA (29), Germany (9), and Japan (5).

It is interesting to note that 20 papers (32%) were authored by reseachers working in industrial labs. This shows clearly the industrial and economic relevance of our field.

There was an increase of papers (6) based on transatlantic collaboration between US researchers and researchers from Finland, France, Israel, Italy, Romania and the United Kingdom. This is a very encouraging trend that is certainly due to increased telecooperation based on the Internet.

The technical program will consist of two full tracks, and occasionally a third parallel session for student papers. Unfortunately, given space considerations, many very good papers could not be accepted. However, the substantial increase in submissions will undoubtedly cause future organizers to consider further changes in conference structure.

In addition to the technical program, this year there are four tutorials, a record twelve workshops over two days, two student sessions with ten papers, numerous demonstrations, and book exhibits. We are also fortunate to have David Sadek from France Telecom and Nino Varile from the European Commission delivering invited addresses. David Sadek will speak about "Rational Agency as the Basis for Natural Dialogue: The ARTIMIS Technology", and Nino Varile will discuss "Current and Future EU Activities in the Field of Language Technologies".

We would like to thank everybody who helped to bring about the richness and high standards of the ACL-EACL'97 conference. We thank all of the authors of submitted papers for choosing ACL-EACL'97 as a forum for the communication of their research results. We have been constantly impressed by the dedication of the Area Chairs and other Program Committee Members, who worked very hard to get the best possible job done. Special praise goes to our workshop coordinator, Harald Trost from the Austrian Research Institute for AI, who ably arranged the post-conference workshops, and to Megumi Kameyama from SRI International, who put together an excellent slate of tutorial speakers. We are very grateful to Pamela Jordan and Johan Bos, who organized two most interesting student sesssions.

We would like to express our gratitude to Kathy McKeown and Priscilla Rasmussen, who provided support and guidance for our work. In addition, we are most appreciative of the outstanding job performed by María Felisa Verdejo, the local arrangements chair, and her Local Organizing Committee in Madrid.

We would like to thank the German Research Center for Artificial Intelligence (DFKI) for providing internal financial support for both administrative and technical tasks related to the conference.

Ira Smith from OGI, who handled the electronic submissions, designed and maintained the submission database, Jochen Müller from DFKI, who designed the interactive Web form for reviewing, maintained the review database and formatted the front pages, Gabi Jacquinot, who maintained the PC database and helped with the Proceedings, and Vivienne Wicks, who arranged a very pleasant PC meeting in Brighton, deserve special mention for the excellent support that they provided.

Welcome to the 1997 ACL-EACL joint meeting at Madrid!

Philip R. Cohen and Wolfgang Wahlster
Program Co-Chairs
May 1997

PC Chairs

Philip R. Cohen — **Oregon Graduate Institute, USA**
Wolfgang Wahlster — **DFKI Saarbrücken, Germany**

Area Chairs

Elisabeth André	DFKI Saarbrücken, Germany (Uses of Language Processing)
Eugene Charniak	Brown Univ., USA (Statistical Language Processing)
Lauri Karttunen	Xerox PARC, USA (Morphology, Lexicon, and Finite State Technologies)
Donia Scott	Univ. of Brighton, United Kingdom (Semantics, Pragmatics and Discourse)
Koenraad de Smedt	Univ. of Bergen, Norway (Grammar and Formalisms for Parsing and Tactical Generation)

PC Members

Anne	Abeille	Univ. of Paris 7, France
Steven	Abney	Univ. of Tübingen, Germany
Doug	Appelt	SRI International, USA
Susan	Armstrong	ISSCO, Switzerland
Hans-Ulrich	Block	Siemens, Germany
Branimir	Boguraev	Apple Computer, USA
Gosse	Bouma	Univ. of Groningen, The Netherlands
Ted	Briscoe	Cambridge Univ., United Kingdom
Sandra	Carberry	Univ. of Delaware, USA
Kenneth Ward	Church	AT&T Research, USA
Robert	Dale	Microsoft Research Institute, Australia
Mary	Dalrymple	Xerox PARC, USA
Giacomo	Ferrari	Univ. of Turin, Italy
Gregory	Grefenstette	Rank Xerox Research Centre, France
Takao	Gunji	Osaka Univ., Japan
Lynette	Hirschman	Mitre, USA
Hitoshi	Iida	ATR Research Laboratories, Japan
David	Israel	SRI International, USA
Michael	Johnston	Oregon Graduate Institute, USA
Ron	Kaplan	Xerox PARC, USA
Genichiro	Kikui	NTT, Japan
Kimmo	Koskenniemi	Univ. of Helsinki, Finland
Alex	Lascarides	Univ. of Edinburgh, United Kingdom
David	Magerman	Renaissance Technologies, USA
Johanna	Moore	Univ. of Pittsburgh, USA
Renato	de Mori	Univ. of Avignon, France
Günter	Neumann	DFKI Saarbrücken, Germany
Hermann	Ney	Aachen Univ. of Technology, Germany
Gertjan	van Noord	Univ. of Groningen, The Netherlands
Kemal	Oflazer	Bilkent Univ., Turkey
James	Pustejovski	Brandeis Univ., USA
Owen	Rambow	Cogentex, USA
Ehud	Reiter	Univ. of Aberdeen, United Kingdom
Uwe	Reyle	Univ. of Stuttgart, Germany

Ellen	Riloff	Univ. of Utah, USA
Horacio	Rodriguez	Univ. Politecnica de Catalunya, Spain
Mats	Rooth	Univ. of Stuttgart, Germany
David	Sadek	France Telecom, France
Patrick	Saint-Dizier	Univ. of Toulouse, France
Yves	Schabes	Mitsubishi Electric Research Lab., USA
Jörg	Schütz	IAI, Germany
Hinrich	Schütze	Xerox PARC, USA
Stuart	Shieber	Harvard Univ., USA
Richard	Sproat	Bell Laboratories, USA
Mark	Steedman	Univ. of Pennsylvania, USA
Andreas	Stolcke	SRI International, USA
Louis	des Tombe	Univ. of Utrecht, The Netherlands
Harald	Trost	Austrian Research Institute for AI, Austria
Marilyn	Walker	AT&T Research, USA
Leo	Wanner	ISI/USC, USA
Peter	Whitelock	Sharp Labs of Europe, United Kingdom
Dekai	Wu	Hong Kong Univ. of Science and Technology, Hong Kong
Steve	Young	Cambridge Univ., United Kingdom

Additional Reviewers

Jan	Alexandersson	Leonardo	Lesmo
Liliana	Ardissono	Hector	Levesque
Etienne	Barnard	Vincenzo	Lombardo
Tilman	Becker	Kavi	Mahesh
Thomas	Bilgram	Elisabeth	Maier
Philippe	Bretier	Henri	Meloni
Thomas	Brückner	Bernard	Merialdo
Bob	Carpenter	Dale	Miller
Jim	Cowie	Larry	Moss
Nils	Dahlbaeck	Stephan	Müller
Laurence	Danlos	Christine	Nanz
Mark	Davis	Mark-Jan	Nederhof
Andreas	Eisele	Sergei	Nirenburg
Marc	El-Beze	Karel	Oliva
Tim	Fernando	Franck	Panaget
Mark	Gawron	Stephen	Pulman
Karin	Harbusch	Norbert	Reithinger
Stephen	Helmreich	Tobias	Ruland
Mark	Johnson	Christer	Samuelsson
Megumi	Kameyama	Giorgio	Satta
Andy	Kehler	Dario	Sestero
Bill	Keller	Richmond H.	Thomason
Esther	König	Evelyne	Viegas
Steven	Krauwer	Theo	Vosse
Hans-Ulrich	Krieger	Nigel	Ward
William	Ladusaw	Frank	Wedekind
Mark	Lauer	Remi	Zajac

ACL-EACL 1997 PROGRAM

SUNDAY EVENING, 6 JULY

Tutorial Registration **7:00PM-9:00PM**
Edificio de Humanidades, UNED, c/ Senda del Rey s/n
Tutorial Reception **7:00PM-9:00PM**
Edificio de Humanidades, UNED, c/ Senda del Rey s/n

MONDAY, 7 JULY

Tutorial Registration **8:00AM-1:00PM**
Morning Tutorials:
Machine Learning of Natural Language **9:30AM-1:00PM**
David Powers
Information Retrieval from a Linguist's Perspective **9:30AM-1:00PM**
Sebastian Goeser and Gerda Ruge
Lunch **1:00PM-3:00PM**
Afternoon Tutorials:
Maximum Entropy Modeling for Natural Language **3:00PM-7:00PM**
Eric Sven Ristad
Logical Approaches to Syntactic Theories **3:00PM-7:00PM**
James Rogers and Thomas Cornell
Conference Registration **3:00PM-8:00PM**
Edificio de Humanidades, UNED, c/Senda del Rey s/n
Conference Reception **7:00PM-10:00PM**
Patio del Colegio Mayor Santa Teresa (nearby registration)

TUESDAY, 8 JULY
(Talks Are 30 Minutes)

Registration **8:00AM - 2:00PM & 3:00PM - 7:00PM**

Parsing **9:00AM - 10:30AM**
Interleaving Universal Principles and Relational Constraints over Typed Feature Logic
Thilo Götz and Detmar Meurers
Fast Context-Free Parsing Requires Fast Boolean Matrix Multiplication
Lillian Lee
Three Generative, Lexicalised Models for Statistical Parsing
Michael Collins
Text Classification **9:00AM - 10:30AM**
Expansion of Multi-word Terms for Indexing and Retrieval Using Morphology and Syntax
Christian Jacquemin, Judith L. Klavans, and Evelyne Tzoukermann
Automatic Detection of Text Genre
Brett Kessler, Geoffrey Nunberg, and Hinrich Schütze
Document Classification Using A Finite Mixture Model
Hang Li and Kenji Yamanishi
Break **10:30AM-11:00AM**
Word Sense Disambiguation **11:00AM - 1:00 PM**
Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation
German Rigau, Jordi Atserias and Eneko Agirre
Similarity-Based Methods for Word Sense Disambiguation
Ido Dagan, Lillian Lee, and Fernando Pereira
Using Syntactic Dependency as Local Context to Resolve Word Sense Ambiguity
Dekang Lin
Homonymy and Polysemy in Information Retrieval
Robert Krovetz
Discourse **11:00AM - 1:00 PM**
Learning Features that Predict Cue Usage
Barbara Di Eugenio, Johanna D. Moore and Massimo Paolucci
Expectations in Incremental Discourse Processing
Dan Cristea and Bonnie Webber
The Rhetorical Parsing of Unrestricted Natural Language Texts
Daniel Marcu
Centered Segmentation: Scaling up the Centering Model to Global Discourse Structure
Udo Hahn and Michael Strube
Lunch **1:00PM - 3:00PM**
Practical Aspects of Machine Translation **3:00PM - 4:30PM**
Probing the Lexicon in Evaluating Commercial MT Systems
Martin Volk
Ambiguity Resolution for Machine Translation of Telegraphic Messages
Young-Suk Lee, Clifford Weinstein, Stephanie Seneff and Dinesh Tummala
Machine Transliteration
Kevin Knight and Jonathan Graehl
Lexicon **3:00PM - 4:30 PM**
Intergrating Symbolic and Statistical Representations: The Lexicon Pragmatics Interface
Ann Copestake and Alex Lascarides
Negative Polarity Licensing at the Syntax-Semantics Interface
John Fry
Deriving Verbal and Compositonal Lexical Aspect for NLP Applications
Bonnie J. Dorr and Mari Broman Olsen
Break **4:30PM - 5:00PM**
Statistics and Meaning **5:00PM - 7:00PM**
A DOP Model for Semantic Interpretation
Remko Bonnema, Rens Bod and Remko Scha
Fertility Models for Statistical Natural Language Understanding
Stephen Della Pietra, Mark Epstein, Salim Roukos and Todd Ward
Predicting the Semantic Orientation of Adjectives
Vasileios Hatzivassiloglou and Kathleen R. McKeown
Independence Assumptions Considered Harmful
Alexander Franz
Generation **5:00PM - 7:00PM**
Planning Reference Choices for Argumentative Texts
Xiaorong Huang
Sentence Planning as Description Using Tree Adjoining Grammar
Matthew Stone and Christine Doran
An Algorithm for Generating Referential Descriptions with Flexible Interfaces
Helmut Horacek
Applying Explanation-based Learning to Control and Speeding-up Natural Language Generation
Günter Neumann

WEDNESDAY, 9 JULY
(Talks Are 30 Minutes Except Invited Talk; Student Session 1 & 2 talks are under 20 minutes)

INVITED TALK **9:00AM - 10:15AM**
Rational Agency as the Basis for Natural Dialogue: The ARTIMIS Technology
David Sadek

Break **10:15AM - 11:00AM**
Morphological Disambiguation **11:00AM - 1:00PM**
Morphological Disambiguation by Voting Constraints
Kemal Oflazer and Gökhan Tür
Mistake-Driven Mixture of Hierarchical Tag Context Trees
Masahiko Haruno and Yuji Matsumoto
A Flexible POS Tagger Using an Automatically Acquired Language Model
Lluís Màrquez and Lluís Padró
Comparing a Linguistic and a Stochastic Tagger
Christer Samuelsson and Atro Voutilainen

ACL-EACL 1997 PROGRAM

Spoken and Multimodal Interaction **11:00AM - 1:00PM**

Intonational Boundaries, Speech Repairs, and Discourse Markers: Modeling Spoken Dialog
Peter A. Heeman and James F. Allen

Tracking Initiative in Collaborative Dialogue Interactions
Jennifer Chu-Carroll and Michael K. Brown

PARADISE: A Framework for Evaluating Spoken Dialogue Agents
Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella

Unification-based Multimodal Integration
Michael Johnston, Philip R. Cohen, David McGee, Sharon L. Oviatt, James A. Pittman, and Ira Smith

Student Session 1 **11:00AM - 12:30PM**

A Structured Language Model
Ciprian Chelba

Incorporating Context Information for the Extraction of Terms
Katerina T. Frantzi

Knowledge Acquisition from Texts: Using an Automatic Clustering Method Based on Noun-Modifier Relationship
Houssem Assadi

Choosing the Word Most Typical in Context Using a Lexical Co-occurrence Network
Philip Edmonds

Improving Translation through Contextual Information
Maite Taboada

Lunch **1:00PM - 3:00PM**

Statistical Machine Translation **3:00PM - 4:30PM**

A DP-based Search Using Monotone Alignments in Statistical Translation
Christoph Tillmann, Stephan Vogel, Hermann Ney and Alex Zubiaga

An Alignment Method for Noisy Parallel Corpora Based on Image Processing Techniques
Jason S. Chang and Mathis H. Chen

A Portable Algorithm for Mapping Bitext Correspondence
I. Dan Melamed

Finite State Technologies **3:00PM - 4:30PM**

Efficient Generation in Primitive Optimality Theory
Jason Eisner

A Trainable Rule-Based Algorithm for Word Segmentation
David D. Palmer

Compiling Regular Formalisms with Rule Features into Finite-State Automata
George Anton Kiraz

Student Session 2 **3:00PM - 4:30PM**

Generative Power of CCGs with Generalized TAGs
Nobo Komagata

Representing Paraphrases Using Synchronous Tree Adjoining Grammars
Mark Dras

Contrastive accent in a data-to-speech system
Mariet Theune

Towards resolution of bridging descriptions
Renata Vieira and Simone Teufel

Compositional Semantics of German Verb Prefixes
Maria Wolters

Break **4:30PM - 5:00PM**

Syntax and Morphology **5:00PM - 6:30PM**

The Complexity of Recognition of Linguistically Adequate Dependency Grammars
Peter Neuhaus and Norbert Bröker

Maximal Incrementality in Linear Categorial Deduction
Mark Hepple

Automatic Extraction of Aspectual Information from a Monolingual Corpus
Akira Oishi and Yuji Matsumoto

Machine Translation and Language Modeling **5:00PM - 7:00PM**

A Comparison of Head Transducers and Transfer for a Limited Domain Translation Application
Hiyan Alshawi, Adam L. Buchsbaum, and Fei Xia

Decoding Algorithm in Statistical Machine Translation
Ye-Yi Wang and Alex Waibel

A Model of Lexical Attraction and Repulsion
Doug Beeferman, Adam Berger and John Lafferty

Hierarchical Non-Emitting Markov Models
Eric Sven Ristad and Robert G. Thomas

THURSDAY, 10 JULY
(Talks Are 30 Minutes, Except Invited Talk)

INVITED TALK **9:00AM - 10:15AM**

Current and Future EU Activities in the Field of Language Technologies
Nino Varile

Break **10:15AM - 11:00AM**

Underspecification and Parallelism **11:00AM - 1:00PM**

Efficient Construction of Underspecified Semantics under Massive Ambiguity
Jochen Dörre

A Theory of Parallelism and the Case of VP Ellipsis
Jerry R. Hobbs and Andrew Kehler

On Interpreting F-structures as UDRSs
Josef van Genabith and Richard Crouch

A Uniform Approach to Underspecification and Parallelism
Joachim Niehren, Manfred Pinkal, and Peter Ruhrberg

Learning, Analogy, and Evolution **11:00AM - 1:00PM**

Co-Evolution of Language and of the Language Acquisition Device
Ted Briscoe

Paradigmatic Cascades: A Linguistically Sound Model of Pronunciation by Analogy
François Yvon

Memory-Based Learning: Using Similarity for Smoothing
Jakub Zavrel and Walter Daelemans

String Transformation Learning
Giorgio Satta and John C. Henderson

Lunch **1:00PM - 3:00PM**

ACL Business Meeting **3:00PM - 4:30PM**

Break **4:30PM - 5:00PM**

Finite State Approximation **5:00PM - 6:30 PM**

Approximating Context-Free Grammars with a Finite-State Calculus
Edmund Grimley Evans

Finite State Transducers Approximating Hidden Markov Models
André Kempe

Representing Constraints with Automata
Frank Morawietz and Tom Cornell

Machine Translation **5:00PM - 6:30PM**

Retrieving Collocations by Co-Occurences and Word Order Constraints
Sayori Shimohata, Toshiyuki Sugio, and Junji Nagata

Learning Parse and Translation Decisions from Examples with Rich Context
Ulf Hermjakob and Raymond J. Mooney

A Word-to-Word Model of Translational Equivalence
I. Dan Melamed

TUTORIALS

Monday 7 July 1997, 9:30AM--6:30PM

Machine Learning of Natural Language **9:30AM--1:00PM**

David Powers, The Flinders University of South Australia

Over the last 30 years, issues relating to how language can be learned have shaped fields as diverse as linguistics, automata theory, and psycholinguistics. More recently, machine learning and neural nets have found linguistic problems a natural target for exploration and demonstration of techniques. Conversely, computational linguistics, natural language processing and speech technology are all actively looking for ways in which learning can be put to practical use in specific applications. Indeed, the increasing availability of large corpora and treebanks promises that the recent explosion in interest will continue.

This tutorial will provide the participant with an understanding of the perspectives on language learning adopted by the different fields, will explore the nature and significance of the various theoretical results, and will characterize and explain the basic machine learning paradigms and algorithms which have been used for language learning. Recent work will be characterized in terms of this framework, with more detailed case studies being drawn from two different paradigms. The focus here will be on developing an intuition as to how and why the various techniques work.

David Powers is co-author of the Springer monograph 'Machine Learning of Natural Language' and is currently President of SIGNLL, the ACL Special Interest Group on Natural Language Learning. This tutorial will provide background useful to those contemplating attending SIGNLL's two day CoNLL workshop.

Information Retrieval from a Linguist's Perspective **9:30AM--1:00PM**

Sebastian Göser, IBM Germany, Gerda Ruge, Technical University of Munich

Information Retrieval (IR) is a dynamically growing field of research and development activities which offers many substantial job opportunities to linguists. The purpose of this tutorial is to introduce computational or theoretical linguists to IR, thus enabling them to make reasonable decisions on linguistic issues in an IR context. The tutorial will give an overview of all major issues in IR research and development that are relevant to a linguist in that context.

An important focus will be the effectiveness of linguistic approaches to IR: Which methods are provably effective, and which ones, in spite of arguments to the contrary, are not? The literature on linguistic IR experiments reports many conflicting results. The IR community is split into those who believe language technology can improve retrieval results and those who do not. The entire IR machinery necessary to understand this situation, in terms of retrieval models, evaluation, system architectures, applications, linguistic representation etc. will be focused in this tutorial.

Maximum Entropy Modeling for Natural Language **3:00PM--6:30PM**

Eric Sven Ristad, Princeton University

The maximum entropy framework is a powerful method for building statistical models of natural language. It is expressive, allowing modelers to easily represent their special insights into the data generating machinery. It is statistically efficient, because it models the intersection of complex events without increasing the number of parameters or fragmenting the training data. And it provides strong models, models that can outperform their traditional variants with less tweaking. This tutorial explains how to build maximum entropy models for natural language applications such as information retrieval and speech recognition. We review the maximum entropy framework, explore the art of effective feature design, and show how to implement models using the instructor's publicly available Maximum Entropy Modeling Toolkit.

Logical Approaches to Syntactic Theories **3:00PM--6:30PM**

James Rogers, University of Central Florida, Thomas Cornell, SFB 340, University of Tübingen

The trend, over the last ten or fifteen years, has been towards specifying syntactic structures in terms of constraints on their form rather than via mechanisms for generating them. This leads naturally to a fully declarative approach in which sets of syntactic structures are treated as sets of ordinary mathematical models and theories of syntax are defined by systems of logical axioms. This model-theoretic approach allows existing tools of mathematical logic to be applied to formal issues in syntax. In addition to the obvious applications to questions of consistency and independence of sets of constraints, results have been obtained establishing the complexity of constraint based theories, logical axiomatizations have been employed as a common framework in which to compare strongly dissimilar generative systems, and automata-based proof techniques have begun to be explored as novel approaches to processing languages axiomatized in this way. This tutorial will explore the foundations of this approach, the results it has yielded so far, and the potential it holds. The presentation will presume only a passing familiarity with basic formal logic and traditional theories of syntax (e.g., GB and GPSG).

TABLE OF CONTENTS

Interleaving Universal Principles and Relational Constraints over Typed Feature Logic
Thilo Götz and Detmar Meurers 1

Fast Context-Free Parsing Requires Fast Boolean Matrix Multiplication
Lillian Lee 9

Three Generative, Lexicalised Models for Statistical Parsing
Michael Collins 16

Expansion of Multi-Word Terms for Indexing and Retrieval Using Morphology and Syntax
Christian Jacquemin, Judith L. Klavans, Evelyne Tzoukermann 24

Automatic Detection of Text Genre
Brett Kessler, Geoffrey Nunberg, and Hinrich Schütze 32

Document Classification Using a Finite Mixture Model
Hang Li, Kenji Yamanishi 39

Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation
German Rigau, Jordi Atserias, Eneko Agirre 48

Similarity-Based Methods for Word Sense Disambiguation
Ido Dagan, Lillian Lee, Fernando Pereira 56

Using Syntactic Dependency as Local Context to Resolve Word Sense Ambiguity
Dekang Lin 64

Homonymy and Polysemy in Information Retrieval
Robert Krovetz 72

Learning Features that Predict Cue Usage
Barbara Di Eugenio, Johanna D. Moore, Massimo Paolucci 80

Expectations in Incremental Discourse Processing
Dan Cristea, Bonnie Webber 88

The Rhetorical Parsing of Unrestricted Natural Language Texts
Daniel Marcu 96

Centered Segmentation: Scaling up the Centering Model to Global Discourse Structure
Udo Hahn, Michael Strube 104

Probing the Lexicon in Evaluating Commercial MT Systems
Martin Volk 112

Ambiguity Resolution for Machine Translation of Telegraphic Messages
Young-Suk Lee, Clifford Weinstein, Stephanie Seneff, Dinesh Tummala 120

Machine Transliteration
Kevin Knight, Jonathan Graehl 128

Intergrating Symbolic and Statistical Representations: The Lexicon Pragmatics Interface
Ann Copestake, Alex Lascarides 136

Negative Polarity Licensing at the Syntax-Semantics Interface
John Fry 144

Deriving Verbal and Compositonal Lexical Aspect for NLP Applications
Bonnie J. Dorr, Mari Broman Olsen 151

A DOP Model for Semantic Interpretation
Remko Bonnema, Rens Bod, Remko Scha 159

Fertility Models for Statistical Natural Language Understanding
Stephen Della Pietra, Mark Epstein, Salim Roukos, Todd Ward 168

Predicting the Semantic Orientation of Adjectives
Vasileios Hatzivassiloglou, Kathleen R. McKeown 174

Independence Assumptions Considered Harmful
Alexander Franz 182

Planning Reference Choices for Argumentative Texts
Xiaorong Huang 190

Sentence Planning as Description Using Tree Adjoining Grammar
Matthew Stone, Christine Doran 198

An Algorithm for Generating Referential Descriptions with Flexible Interfaces
Helmut Horacek 206

Applying Explanation-based Learning to Control and Speeding-up Natural Language Generation
Günter Neumann 214

Morphological Disambiguation by Voting Constraints
Kemal Oflazer and Gökhan Tür 222

Mistake-Driven Mixture of Hierarchical Tag Context Trees
Masahiko Haruno, Yuji Matsumoto 230

A Flexible POS Tagger Using an Automatically Acquired Language Model
Lluís Màrquez and Lluís Padró 238

Comparing a Linguistic and a Stochastic Tagger
Christer Samuelsson, Atro Voutilainen 246

Intonational Boundaries, Speech Repairs, and Discourse Markers: Modeling Spoken Dialog
Peter A. Heeman, James F. Allen 254

Tracking Initiative in Collaborative Dialogue Interactions
Jennifer Chu-Carroll and Michael K. Brown 262

PARADISE: A Framework for Evaluating Spoken Dialogue Agents
Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, Alicia Abella 271

Unification-based Multimodal Integration
Michael Johnston, Philip R. Cohen, David McGee, Sharon L. Oviatt, James A. Pittman, and Ira Smith 281

A DP-based Search Using Monotone Alignments in Statistical Translation
Christoph Tillmann, Stephan Vogel, Hermann Ney and Alex Zubiaga 289

An Alignment Method for Noisy Parallel Corpora based on Image Processing Techniques
Jason S. Chang, Mathis H. Chen 297

A Portable Algorithm for Mapping Bitext Correspondence
I. Dan Melamed 305

Efficient Generation in Primitive Optimality Theory
Jason Eisner 313

A Trainable Rule-Based Algorithm for Word Segmentation
David D. Palmer 321

Compiling Regular Formalisms with Rule Features into Finite-State Automata
George Anton Kiraz 329

The Complexity of Recognition of Linguistically Adequate Dependency Grammars
Peter Neuhaus and Norbert Bröker 337

Maximal Incrementality in Linear Categorial Deduction
Mark Hepple 344

Automatic Extraction of Aspectual Information from a Monolingual Corpus
Akira Oishi, Yuji Matsumoto 352

A Comparison of Head Transducers and Transfer for a Limited Domain Translation Application
Hiyan Alshawi, Adam L. Buchsbaum, Fei Xia 360

Decoding Algorithm in Statistical Machine Translation
Ye-Yi Wang, Alex Waibel 366

A Model of Lexical Attraction and Repulsion
Doug Beeferman, Adam Berger, John Lafferty 373

Hierarchical Non-Emitting Markov Models
Eric Sven Ristad, Robert G. Thomas 381

Efficient Construction of Underspecified Semantics under Massive Ambiguity
Jochen Dörre 386

A Theory of Parallelism and the Case of VP Ellipsis
Jerry R. Hobbs, Andrew Kehler 394

On Interpreting F-Structures as UDRSs
Josef van Genabith, Richard Crouch 402

A Uniform Approach to Underspecification and Parallelism
Joachim Niehren, Manfred Pinkal, Peter Ruhrberg 410

Co-Evolution of Language and of the Language Acquisition Device
Ted Briscoe 418

Paradigmatic Cascades: A Linguistically Sound Model of Pronunciation by Analogy
François Yvon 428

Memory-Based Learning: Using Similarity for Smoothing
Jakub Zavrel, Walter Daelemans 436

String Transformation Learning
Giorgio Satta, John C. Henderson 444

Approximating Context-Free Grammars with a Finite-State Calculus
Edmund Grimley Evans 452

Finite State Transducers Approximating Hidden Markov Models
André Kempe 460

Representing Constraints with Automata
Frank Morawietz and Tom Cornell 468

Retrieving Collocations by Co-Occurrences and Word Order Constraints
Sayori Shimohata, Toshiyuki Sugio, Junji Nagata 476

Learning Parse and Translation Decisions from Examples with Rich Context
Ulf Hermjakob, Raymond J. Mooney 482

A Word-to-Word Model of Translational Equivalence
I. Dan Melamed 490

STUDENT SESSION TABLE OF CONTENTS

A Structured Language Model
Ciprian Chelba 498

Incorporating Context Information for the Extraction of Terms
Katerina T. Frantzi 501

Knowledge Acquisition from Texts: Using an Automatic Clustering Method Based on Noun-Modifier Relationship
Houssem Assadi 504

Choosing the Word Most Typical in Context Using a Lexical Co-occurrence Network
Philip Edmonds 507

Improving Translation through Contextual Information
Maite Taboada 510

Generative Power of CCGs with Generalized Type-Raised Categories
Nobo Komagata 513

Representing Paraphrases Using Synchronous TAGs
Mark Dras 516

Contrastive accent in a data-to-speech system
Mariet Theune 519

Towards resolution of bridging descriptions
Renata Vieira, Simone Teufel 522

Compositional Semantics of German Prefix Verbs
Maria Wolters 525

PREFACE TO THE STUDENT SESSION PAPERS

These proceedings include the extended abstracts accepted for presentation at the Student Session of the *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics.* The goal of the Student Session is to provide a forum for student members to present work in progress, rather than completed work, and to receive feedback from other members of the computational linguistics community, particularly senior researchers. The response to the ACL Student Sessions held during the previous years was very positive. The student authors consistently report that they find the Student Sessions valuable, and answers to questionnaires filled out by ACL members (most recently in 1995) indicate that the audiences find the sessions interesting and of high quality.

In previous years, the ACL Student Session had moved from paper presentations to poster presentations while the EACL Student Session had continued with paper presentations. Although this year's Call for Papers advertised a Student poster session, the Program Co-Chairs encouraged us to switch to paper presentations. We thank Wolfgang Wahlster and Phil Cohen for this opportunity to have paper presentations once again.

Forty-two papers were submitted to the Student Session and we accepted ten of these. We thank all the authors for their submissions, and hope that the reviews provided constructive criticism and encouraged them in their research.

We also thank all of the reviewers for the time they spent doing careful reviews of the submissions. Everyone was most cooperative and forgiving as we negotiated matches between reviewers and papers.

Student members of the ACL 1997 Student Session Program Committee: Paul Buitelaar *Brandeis University*; Alastair Butler *University of East Anglia*; Bekki Daisuke *University of Tokyo*. Mariana Damova *University of Stuttgart*; Eric Fosler *U.C. Berkeley*; Rob Koeling, *University of Groningen*; Mark Lee *University of Sheffield*; Maria Milosavljevic *Macquarie University*; Ted Pedersen *Southern Methodist University*; Carolyn Rosé *Carnegie Mellon University*; Bilge Say *Bilkent University*; Michael Schiehlen *University of Stuttgart*; Hadar Shemtov *Stanford University*; B. Srinivas *University of Pennsylvania*; Kjetil Strand *University of Oslo*; David Tugwell *University of Edinburgh*; Peter Vanderheyden *University of Waterloo.*

Non-student members of the Student Session Program Committee: W. Scott Bennett *Logos Corp.*; Martin Emele *University of Stuttgart*; Ted Gibson *Massachusetts Institute of Technology*; Paola Merlo *University of Geneva*; Marie Meteer *BBN*; Susan McRoy *University of Wisconsin-Milwaukee*; Massimo Poesio *University of Edinburgh*; Craige Roberts *Ohio State University*; Patrick Saint-Dizier *Institut de Recherche en Informatique Toulouse*; Koichi Takeda *Tokyo Research Laboratory, IBM Japan*; and Gertjan van Noord *University of Groningen.*

We are grateful too to those who provided us with special reviews. The additional students we wish to thank are: Kathryn Baker, *Carnegie Mellon University* and Matthew Stone, *University of Pennsylvania.* The additional non-student reviewers included: Bianka Buschbeck, *University of Stuttgart*; Nancy Green, *Carnegie Mellon University*; and Mark-Jan Nederhof, *University of Groningen.*

Last year's Student Session chairs, Mettina Veenstra, *University of Groningen* and Christine Doran, *University of Pennsylvania* were extremely generous with their advice. They provided us with a timeline and other invaluable advice on how to conduct a student session. Finally, we wish to thank the Program Co-Chairs, the ACL Executive Committee, Rich Thomason *University of Pittsburgh* and Barbara Di Eugenio *University of Pittsburgh* for their general advice and guidance.

Pamela Jordan, *University of Pittsburgh*
Johan Bos, *University of the Saarland*
Student Session Co-Chairs

AUTHOR INDEX

Abella, Alicia 271
Agirre, Eneko 48
Allen, James F. 254
Alshawi, Hiyan 360
Atserias, Jordi 48
Beeferman, Doug 373
Berger, Adam 373
Bod, Rens 159
Bonnema, Remko 159
Briscoe, Ted 418
Bröker, Norbert 337
Brown, Michael K. 262
Buchsbaum, Adam L. 360
Chang, Jason S. 297
Chen, Mathis H. 297
Chu-Carroll, Jennifer 262
Cohen, Philip R. 281
Collins, Michael 16
Copestake, Ann 136
Cornell, Tom 468
Cristea, Dan 88
Crouch, Richard 402
Daelemans, Walter 436
Dagan, Ido 56
Della Pietra, Stephen 168
Di Eugenio, Barbara 80
Dörre, Jochen 386
Doran, Christine 198
Dorr, Bonnie J. 151
Eisner, Jason 313
Epstein, Mark 168
Franz, Alexander 182
Fry, John 144
Götz, Thilo 1
Graehl, Jonathan 128
Grimley Evans, Edmund 452
Hahn, Udo 104
Haruno, Masahiko 230
Hatzivassiloglou, Vasileios 174
Heeman, Peter A. 254
Henderson, John C. 444
Hepple, Mark 344
Hermjakob, Ulf 482
Hobbs, Jerry R. 394
Horacek, Helmut 206
Huang, Xiaorong 190
Jacquemin, Christian 24
Johnston, Michael 281
Kamm, Candace A. 271
Kehler, Andrew 394
Kempe, André 460
Kessler, Brett 32
Kiraz, George Anton 329
Klavans, Judith L. 24
Knight, Kevin 128
Krovetz, Robert 72
Lafferty, John 373
Lascarides, Alex 136
Lee, Lillian 9, 56
Lee, Young-Suk 120
Li, Hang 39
Lin, Dekang 64
Litman, Diane J. 271
Marcu, Daniel 96
Màrquez, Lluís 238
Matsumoto, Yuji 230, 352
McGee, David 281
McKeown, Kathleen R. 174
Melamed, I. Dan 305, 490
Meurers, Detmar 1
Mooney, Raymond J. 482
Moore, Johanna D. 80
Morawietz, Frank 468
Nagata, Junji 476
Neuhaus, Peter 337
Neumann, Günther 214
Ney, Hermann 289
Niehren, Joachim 410
Nunberg, Geoffrey 32
Oflazer, Kemal 222
Oishi, Akira 352
Olsen, Mari Broman 151
Oviatt, Sharon L. 281
Padró, Lluís 238
Palmer, David D. 321
Paolucci, Massimo 80
Pereira, Fernando 56
Pinkal, Manfred 410
Pittman, James A. 281
Rigau, German 48

Ristad, Eric Sven 381
Roukos, Salim 168
Ruhrberg, Peter 410
Samuelsson, Christer 246
Satta, Giorgio 444
Scha, Remko 159
Schütze, Hinrich 32
Seneff, Stephanie 120
Shimohata, Sayori 476
Smith, Ira 281
Stone, Matthew 198
Strube, Michael 104
Thomas, Robert G. 381
Tillmann, Christoph 289
Toshiyuki, Sugio 476
Tür, Gökhan 222
Tummala, Dinesh 120
Tzoukermann, Evelyne 24
van Genabith, Josef 402
Vogel, Stephan 289
Volk, Martin 112
Voutilainen, Atro 246
Waibel, Alex 366
Walker, Marilyn A. 271
Wang, Ye-Yi 366
Ward, Todd 168
Webber, Bonnie 88
Weinstein, Clifford 120
Xia, Fei 360
Yamanishi, Kenji 39
Yvon, François 428
Zavrel, Jakub 436
Zubiaga, Alex 289

STUDENT AUTHOR INDEX

Assadi, Houssem 504
Chelba, Ciprian 498
Dras, Mark 516
Edmonds, Philip 507
Frantzi, Katerina T. 501
Komagata, Nobo 513
Teufel, Simone 522
Theune, Mariet 519
Toboada, Maite 510
Vieira, Renata 522
Wolters, Maria 525