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
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 truly 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 researchers 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".

III
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 sessions.

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**

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

**PC Members**

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>Univ. of Paris 7, France</td>
</tr>
<tr>
<td>Steven</td>
<td>Univ. of Tübingen, Germany</td>
</tr>
<tr>
<td>Doug</td>
<td>SRI International, USA</td>
</tr>
<tr>
<td>Susan</td>
<td>ISSCO, Switzerland</td>
</tr>
<tr>
<td>Hans-Ulrich</td>
<td>Siemens, Germany</td>
</tr>
<tr>
<td>Branimir</td>
<td>Apple Computer, USA</td>
</tr>
<tr>
<td>Gosse</td>
<td>Univ. of Groningen, The Netherlands</td>
</tr>
<tr>
<td>Ted</td>
<td>Cambridge Univ., United Kingdom</td>
</tr>
<tr>
<td>Sandra</td>
<td>Univ. of Delaware, USA</td>
</tr>
<tr>
<td>Kenneth Ward</td>
<td>AT&amp;T Research, USA</td>
</tr>
<tr>
<td>Robert</td>
<td>Microsoft Research Institute, Australia</td>
</tr>
<tr>
<td>Mary</td>
<td>Xerox PARC, USA</td>
</tr>
<tr>
<td>Giacomo</td>
<td>Univ. of Turin, Italy</td>
</tr>
<tr>
<td>Gregory</td>
<td>Rank Xerox Research Centre, France</td>
</tr>
<tr>
<td>Takao</td>
<td>Osaka Univ., Japan</td>
</tr>
<tr>
<td>Lynette</td>
<td>Mitre, USA</td>
</tr>
<tr>
<td>Hitoshi</td>
<td>ATR Research Laboratories, Japan</td>
</tr>
<tr>
<td>David</td>
<td>SRI International, USA</td>
</tr>
<tr>
<td>Michael</td>
<td>Oregon Graduate Institute, USA</td>
</tr>
<tr>
<td>Ron</td>
<td>Xerox PARC, USA</td>
</tr>
<tr>
<td>Genichiro</td>
<td>NTT, Japan</td>
</tr>
<tr>
<td>Kimmo</td>
<td>Univ. of Helsinki, Finland</td>
</tr>
<tr>
<td>Alex</td>
<td>Univ. of Edinburgh, United Kingdom</td>
</tr>
<tr>
<td>David</td>
<td>Renaissance Technologies, USA</td>
</tr>
<tr>
<td>Johanna</td>
<td>Univ. of Pittsburgh, USA</td>
</tr>
<tr>
<td>Renato</td>
<td>Univ. of Avignon, France</td>
</tr>
<tr>
<td>Günter</td>
<td>DFKI Saarbrücken, Germany</td>
</tr>
<tr>
<td>Hermann</td>
<td>Aachen Univ. of Technology, Germany</td>
</tr>
<tr>
<td>Gertjan</td>
<td>Univ. of Groningen, The Netherlands</td>
</tr>
<tr>
<td>Kemal</td>
<td>Bilkent Univ., Turkey</td>
</tr>
<tr>
<td>James</td>
<td>Brandeis Univ., USA</td>
</tr>
<tr>
<td>Owen</td>
<td>Cogentex, USA</td>
</tr>
<tr>
<td>Ehud</td>
<td>Univ. of Aberdeen, United Kingdom</td>
</tr>
<tr>
<td>Uwe</td>
<td>Univ. of Stuttgart, Germany</td>
</tr>
</tbody>
</table>
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 Schütze  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
Liliana Ardissono
Etienne Barnard
Tilman Becker
Thomas Bilgram
Philippe Bretier
Thomas Brückner
Bob Carpenter
Jim Cowie
Nils Dahlbaek
Laurence Danlos
Mark Davis
Andreas Eisele
Marc El-Beze
Tim Fernando
Mark Gawron
Karin Harbusch
Stephen Helmreich
Mark Johnson
Megumi Kameyama
Andy Kehler
Bill Keller
Esther Köng
Steven Krauwer
Hans-Ulrich Krieger
William Ladusaw
Mark Lauer
Leonardo Lesmo
Hector Levesque
Vincenzo Lombardo
Kavi Mahesh
Elisabeth Maier
Henri Meloni
Bernard Merialdo
Dale Miller
Larry Moss
Stephan Müller
Christine Nanz
Mark-Jan Nederhof
Sergei Nirenburg
Karel Oliva
Franck Panaget
Stephen Pulman
Norbert Reithinger
Tobias Ruland
Christer Samuelsson
Giorgio Satta
Dario Sestero
Richmond H. Thomason
Evelyne Viegas
Theo Vosse
Nigel Ward
Frank Wedekind
Remi Zajac
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 Goess 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, Lexicalized 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
- 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 - 4:30PM
Integrating 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 Compositional 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
  Vasilisios Hatzipannis 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 NLP
  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 Oilaz 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 Marquèz and Lluís Padró
Comparing a Linguistic and a Stochastic Tagger
  Christer Samuelsson and Ato Voutilainen

VII
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 Neuhäus and Norbert Brokér
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 Dörr
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 Ruhberg
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
Girgio 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 - 6:30PM
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-Occurrences 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

VIII
Machine Learning of Natural Language  
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  
Sebastian Goser, 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  
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  
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 Translation
   Kevin Knight, Jonathan Graehl ........................................................ 128

Integrating 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 Compositional 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
<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicting the Semantic Orientation of Adjectives</td>
<td>Vasileios Hatzivassiloglou, Kathleen R. McKeown</td>
<td>174</td>
</tr>
<tr>
<td>Independence Assumptions Considered Harmful</td>
<td>Alexander Franz</td>
<td>182</td>
</tr>
<tr>
<td>Planning Reference Choices for Argumentative Texts</td>
<td>Xiaorang Huang</td>
<td>190</td>
</tr>
<tr>
<td>Sentence Planning as Description Using Tree Adjoining Grammar</td>
<td>Matthew Stone, Christine Doran</td>
<td>198</td>
</tr>
<tr>
<td>An Algorithm for Generating Referential Descriptions with Flexible Interfaces</td>
<td>Helmut Horacek</td>
<td>206</td>
</tr>
<tr>
<td>Applying Explanation-based Learning to Control and Speeding-up Natural Language Generation</td>
<td>Günter Neumann</td>
<td>214</td>
</tr>
<tr>
<td>Morphological Disambiguation by Voting Constraints</td>
<td>Kemal Oflazer, Gökhan Tür</td>
<td>222</td>
</tr>
<tr>
<td>Mistake-Driven Mixture of Hierarchical Tag Context Trees</td>
<td>Masahiko Haruno, Yuji Matsumoto</td>
<td>230</td>
</tr>
<tr>
<td>A Flexible POS Tagger Using an Automatically Acquired Language Model</td>
<td>Lluís Márquez, Lluís Padró</td>
<td>238</td>
</tr>
<tr>
<td>Comparing a Linguistic and a Stochastic Tagger</td>
<td>Christer Samuelsson, Atro Voutilainen</td>
<td>246</td>
</tr>
<tr>
<td>Intonational Boundaries, Speech Repairs, and Discourse Markers: Modeling Spoken Dialog</td>
<td>Peter A. Heeman, James F. Allen</td>
<td>254</td>
</tr>
<tr>
<td>Tracking Initiative in Collaborative Dialogue Interactions</td>
<td>Jennifer Chu-Carroll and Michael K. Brown</td>
<td>262</td>
</tr>
<tr>
<td>Unification-based Multimodal Integration</td>
<td>Michael Johnston, Philip R. Cohen, David McGee, Sharon L. Oviatt, James A. Pittman, and Ira Smith</td>
<td>281</td>
</tr>
<tr>
<td>A DP-based Search Using Monotone Alignments in Statistical Translation</td>
<td>Christoph Tillmann, Stephan Vogel, Hermann Ney and Alex Zubiaga</td>
<td>289</td>
</tr>
<tr>
<td>An Alignment Method for Noisy Parallel Corpora based on Image Processing Techniques</td>
<td>Jason S. Chang, Mathis H. Chen</td>
<td>297</td>
</tr>
<tr>
<td>A Portable Algorithm for Mapping Bitext Correspondence</td>
<td>I. Dan Melamed</td>
<td>305</td>
</tr>
<tr>
<td>Efficient Generation in Primitive Optimality Theory</td>
<td>Jason Eisner</td>
<td>313</td>
</tr>
<tr>
<td>A Trainable Rule-Based Algorithm for Word Segmentation</td>
<td>David D. Palmer</td>
<td>321</td>
</tr>
<tr>
<td>Compiling Regular Formalisms with Rule Features into Finite-State Automata</td>
<td>George Anton Kiraz</td>
<td>329</td>
</tr>
<tr>
<td>The Complexity of Recognition of Linguistically Adequate Dependency Grammars</td>
<td>Peter Neuhaus and Norbert Bröker</td>
<td>337</td>
</tr>
<tr>
<td>Maximal Incrementality in Linear Categorial Deduction</td>
<td>Mark Hepple</td>
<td>344</td>
</tr>
</tbody>
</table>
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; Craig 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 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 Dorran, 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
<table>
<thead>
<tr>
<th>Author Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abella, Alicia</td>
<td>271</td>
</tr>
<tr>
<td>Agirre, Eneko</td>
<td>48</td>
</tr>
<tr>
<td>Allen, James F.</td>
<td>254</td>
</tr>
<tr>
<td>Alshawi, Hiyan</td>
<td>360</td>
</tr>
<tr>
<td>Atserias, Jordi</td>
<td>48</td>
</tr>
<tr>
<td>Beeferman, Doug</td>
<td>373</td>
</tr>
<tr>
<td>Berger, Adam</td>
<td>373</td>
</tr>
<tr>
<td>Bod, Rens</td>
<td>159</td>
</tr>
<tr>
<td>Bonnema, Remko</td>
<td>159</td>
</tr>
<tr>
<td>Briscoe, Ted</td>
<td>418</td>
</tr>
<tr>
<td>Bröker, Norbert</td>
<td>337</td>
</tr>
<tr>
<td>Brown, Michael K.</td>
<td>262</td>
</tr>
<tr>
<td>Buchsbaum, Adam L.</td>
<td>360</td>
</tr>
<tr>
<td>Chang, Jason S.</td>
<td>297</td>
</tr>
<tr>
<td>Chen, Mathis H.</td>
<td>297</td>
</tr>
<tr>
<td>Chu-Carroll, Jennifer</td>
<td>262</td>
</tr>
<tr>
<td>Cohen, Philip R.</td>
<td>281</td>
</tr>
<tr>
<td>Collins, Michael</td>
<td>16</td>
</tr>
<tr>
<td>Copestake, Ann</td>
<td>136</td>
</tr>
<tr>
<td>Cornell, Tom</td>
<td>468</td>
</tr>
<tr>
<td>Cristea, Dan</td>
<td>88</td>
</tr>
<tr>
<td>Crouch, Richard</td>
<td>402</td>
</tr>
<tr>
<td>Daelemans, Walter</td>
<td>436</td>
</tr>
<tr>
<td>Dagan, Ido</td>
<td>56</td>
</tr>
<tr>
<td>Della Pietra, Stephen</td>
<td>168</td>
</tr>
<tr>
<td>Di Eugenio, Barbara</td>
<td>80</td>
</tr>
<tr>
<td>Dörre, Jochen</td>
<td>386</td>
</tr>
<tr>
<td>Doran, Christine</td>
<td>198</td>
</tr>
<tr>
<td>Dorr, Bonnie J.</td>
<td>151</td>
</tr>
<tr>
<td>Eisner, Jason</td>
<td>313</td>
</tr>
<tr>
<td>Epstein, Mark</td>
<td>168</td>
</tr>
<tr>
<td>Franz, Alexander</td>
<td>182</td>
</tr>
<tr>
<td>Fry, John</td>
<td>144</td>
</tr>
<tr>
<td>Götz, Thilo</td>
<td>1</td>
</tr>
<tr>
<td>Graehl, Jonathan</td>
<td>128</td>
</tr>
<tr>
<td>Grimley Evans, Edmund</td>
<td>452</td>
</tr>
<tr>
<td>Hahn, Udo</td>
<td>104</td>
</tr>
<tr>
<td>Haruno, Masahiko</td>
<td>230</td>
</tr>
<tr>
<td>Hatzivassiloglou, Vasileios</td>
<td>174</td>
</tr>
<tr>
<td>Heeman, Peter A.</td>
<td>254</td>
</tr>
<tr>
<td>Henderson, John C.</td>
<td>444</td>
</tr>
<tr>
<td>Hepple, Mark</td>
<td>344</td>
</tr>
<tr>
<td>Hermjakob, Ulf</td>
<td>482</td>
</tr>
<tr>
<td>Hobbs, Jerry Ulf</td>
<td>394</td>
</tr>
<tr>
<td>Horacek, Helmut</td>
<td>206</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang, Xiaorong</td>
<td>190</td>
</tr>
<tr>
<td>Jacquemin, Christian</td>
<td>24</td>
</tr>
<tr>
<td>Johnston, Michael</td>
<td>281</td>
</tr>
<tr>
<td>Kamm, Candace A.</td>
<td>271</td>
</tr>
<tr>
<td>Kehler, Andrew</td>
<td>394</td>
</tr>
<tr>
<td>Kempe, André</td>
<td>460</td>
</tr>
<tr>
<td>Kessler, Brett</td>
<td>32</td>
</tr>
<tr>
<td>Kiraz, George Anton</td>
<td>329</td>
</tr>
<tr>
<td>Klavans, Judith L.</td>
<td>24</td>
</tr>
<tr>
<td>Knight, Kevin</td>
<td>128</td>
</tr>
<tr>
<td>Krovetz, Robert</td>
<td>72</td>
</tr>
<tr>
<td>Lafferty, John</td>
<td>373</td>
</tr>
<tr>
<td>Lascarides, Alex</td>
<td>136</td>
</tr>
<tr>
<td>Lee, Lillian</td>
<td>9, 56</td>
</tr>
<tr>
<td>Lee, Young-Suk</td>
<td>120</td>
</tr>
<tr>
<td>Li, Hang</td>
<td>39</td>
</tr>
<tr>
<td>Lin, Dekang</td>
<td>64</td>
</tr>
<tr>
<td>Litman, Diane J.</td>
<td>271</td>
</tr>
<tr>
<td>Lascarides, Alex</td>
<td>136</td>
</tr>
<tr>
<td>Lucan, Daniel</td>
<td>96</td>
</tr>
<tr>
<td>Márquez, Lluis</td>
<td>238</td>
</tr>
<tr>
<td>Matsumoto, Yuji</td>
<td>230, 352</td>
</tr>
<tr>
<td>McGee, David</td>
<td>281</td>
</tr>
<tr>
<td>McKeown, Kathleen R.</td>
<td>174</td>
</tr>
<tr>
<td>Melamed, I. Dan</td>
<td>305, 490</td>
</tr>
<tr>
<td>Meurers, Detmar</td>
<td>1</td>
</tr>
<tr>
<td>Mooney, Raymond J.</td>
<td>482</td>
</tr>
<tr>
<td>Moore, Johanna D.</td>
<td>80</td>
</tr>
<tr>
<td>Morawietz, Frank</td>
<td>468</td>
</tr>
<tr>
<td>Neumann, Günther</td>
<td>214</td>
</tr>
<tr>
<td>Néy, Hermann</td>
<td>289</td>
</tr>
<tr>
<td>Niehren, Joachim</td>
<td>410</td>
</tr>
<tr>
<td>Nunberg, Geoffrey</td>
<td>32</td>
</tr>
<tr>
<td>Oflazer, Kemal</td>
<td>222</td>
</tr>
<tr>
<td>Oishi, Akira</td>
<td>352</td>
</tr>
<tr>
<td>Olsen, Mari Broman</td>
<td>151</td>
</tr>
<tr>
<td>Oviatt, Sharon L.</td>
<td>281</td>
</tr>
<tr>
<td>Padró, Lluis</td>
<td>238</td>
</tr>
<tr>
<td>Palmer, David D.</td>
<td>321</td>
</tr>
<tr>
<td>Paolucci, Massimo</td>
<td>80</td>
</tr>
<tr>
<td>Pereira, Fernando</td>
<td>56</td>
</tr>
<tr>
<td>Pinkal, Manfred</td>
<td>410</td>
</tr>
<tr>
<td>Pittman, James A.</td>
<td>281</td>
</tr>
<tr>
<td>Rigau, German</td>
<td>48</td>
</tr>
</tbody>
</table>

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

STUDENT AUTHOR INDEX

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

XVI