ACL-02

40th Annual Meeting of the Association for Computational Linguistics

Jointly hosted by the North American chapter of the ACL

Proceedings of the Conference

General Chair: Pierre Isabelle
Program Co-Chairs: Eugene Charniak and Dekang Lin

7 –12 July 2002
University of Pennsylvania
Philadelphia, Pennsylvania, USA

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

The conference, the workshops and the tutorials are sponsored by the following organizations.
IBM is a registered trademark of International Business Machines Corporation.

SILVER SPONSORS

LDC LINGUISTIC DATA CONSORTIUM

Microsoft® Research

We make the net work. Sun microsystems

2^40 TERAGRAM Linguistic Technologies

BRONZE SPONSORS

AT&T

BASIS TECHNOLOGY

BBN TECHNOLOGIES A Verizon Company

Intelligent Information Systems Institute CORNELL

LexisNexis™

SPECIAL EVENT AND OTHER SPONSORS

gsk GlaxoSmithKline (banquet band)

Xerox Research Centre Europe xrce A Xerox Research & Technology Organization (general support)

IBM® (student business lunch)

Microsoft® Research (student travel)

Penn Engineering™ (reception)

Copyright @2002
The Association for Computational Linguistics
Order copies of this and other ACL publications from:
Morgan Kaufmann Publishers
San Francisco
http://www.mkp.com
orders@mkp.com
1-800-745-7323 (phone)
1-800-814-6418 (fax)
ISBN: 1-55860-883-4

ADMINISTRATION

IRCS Institute for Research in Cognitive Science

Preface: General Chair

This year's meeting is special. First, it the first ACL meeting to be hosted jointly with our North-American chapter, the NAACL. Second, it is also the ACL's 40th anniversary meeting: our association was created 13 June 1962, and its first meeting took place 25-26 August 1963 in Denver.

Our field has expanded quite a bit since then, and so have ACL meetings. This year, our main conference program includes some 65 different refereed papers, covering most active research areas in computational linguistics. And our extended program is complementing that with 19 system demonstrations, 4 tutorials, 12 Student Research papers, and no less than 11 different pre- and post-conference workshops presenting a total of about 150 papers.

My first thanks are addressed to all those who chose to submit their work at ACL-02. Collectively, they spent some 5850 person-days (roughly 25 person-years) just for preparing submissions, final versions and oral presentations! [1] Of course, the underlying research effort is very much larger.

Eugene Charniak and Dekang Lin, our Program Co-Chairs for the main conference skillfully steered their highly competent program committee towards elaborating an enticing and compelling scientific program. Conference highlights include invited talks by Daphne Koller and Peter Norvig, as well as a Best Paper Award.

Some special events have also been attached to the main program of this 40th Anniversary meeting: 1) the new ACL electronic anthology will be presented; and 2) an ACL Lifetime Achievement Award will for the first time be presented to one of our outstanding contributors.

Ronnie Smith, our Demo Chair, and his reviewers Steven Bird and Johan Boyes selected 19 demos that will provide a vivid illustration of what can be achieved with current leading edge technology.

Our Student Research Workshop Co-Chairs, Alexander Koller and Gideon Mann, with some guidance from their faculty advisor Susan McRoy, have set up a student program that meets the highest quality standards. See companion volume for details.

Our Tutorials Chair, Hwee Tou Ng, and his two reviewers, Pamela Jordan and Jakub Zavrel, selected 4 tutorials on important and timely topics, to be presented by highly regarded specialists: Sadaoki Furui, David D. Lewis, Jean-Michel Renders and Michael Strube. See companion volume for details.

Walter Daelemans, our Workshops Chair, elicited great workshop proposals and coached workshop organizers through all the intricacies of ACL-02 organizational requirements. Eleven workshop projects were selected, ranging from popular SIG meetings (SIGDAT, SIGDIAL, SIGPHON-SIGNLL) and hot topics (summarization, NLP in bioinformatics) to more specialized and more speculative topics. Ten of the workshops have their own proceedings.

Schedule constraints are such that final versions of scientific contributions can only become available a few weeks before the conference. Dragomir Radev, our Publication Chair, had the challenging mission to get them all published in time for the conference, both in hardcopy and on CD-ROM. He received the help of several people, including: Dekang Lin, who organized Web-based collection of all electronic versions, David Yarowsky who provided code and advice for producing the CD-ROM,

[1] This is a rough estimate that assumes an average of 7 person-days for preparing each submission and an average of 4 person-days for preparing final versions plus slides.

and Jahna Otterbacher who did most of the CD-ROM development work. Huge thanks to all of them!

Our Webmaster Aurélien Max, did exemplary work setting up and maintaining a convenient, informative and up-to-date ACL-02 Web site (**http://www.acl02.org**).

The ACL has a tradition of keeping registration fees affordable for all, and especially for students. Kathy McCoy, our ACL secretary-treasurer, did a lot of hard work to help us meet that goal without any need to nibble away on the convenience or even on the fun!

A key factor in balancing our budget has been the great skills of Deborah Dahl, our Sponsorships/Exhibits Chair, in successfully eliciting substantial sponsorships. I should also mention Susan McRoy, who secured an NSF grant to help students attend the conference. My most sincere thanks to all ACL-02 sponsors for their generosity!

I would also like to thank Priscilla Rasmussen, the ACL business manager, who provided us with expert guidance and advice about many delicate issues and important details.

Finally, our UPenn hosts deserve our deep gratitude. They have set the finest possible table for the grand scientific meal prepared by our authors. Believe me, "local arrangements" is a lot of work! Martha Palmer has set up and managed an outstanding team to do it all in great style: Trisha Yannuzzi, Laurel Sweeney, Gilbert Kowie, Craig Martell, Yuan Ding, Paul Howard.

So many people contributed to organizing ACL-02 in one way or another that I am sure I forgot to mention some essential contributions here. Thanks to all those who helped!

In closing I would like to thank you all for attending ACL-02, and I wish you a fruitful and pleasant conference.

Pierre Isabelle
May 23, 2002

Preface: Program Committee Co-Chairs

Actually, producing one of these volumes is very simple. First you make sure you have a program co-chair who is willing to do all the work. Then you check that you are working for a general conference chair and executive committee who answer their email quickly so that, even if it takes 10 or so emails to pin down the official abbreviation for the conference, the actual elapsed time is under a week. (It was, just barely, but despite this one snafu they were really great.) Next you round up a great group of very responsible area chairs, who in turn recruit (mostly) responsible reviewers. Then you wait for 258 papers to arrive from 22 countries (130 from the Americas, 53 Asia-Pacific, and 75 Europe-Africa), send them out to the reviewers, get back (most of) the reviews and go to the program committee meeting. This last bit, however requires you to have someone who will agree to host the committee meeting in a town with better flight connections than either Edmonton or Providence. For this we would like to thank Graeme Hirst of the University of Toronto, who did a fabulous job of making the committee meeting a success. You then hope that your fine program committee members not only are responsible but also have great senses of the absurd, so that after a day and a half of tough decision making everyone is still happy and working hard. They had and they did. Then you line up some great invited speakers, and we would like to thank Daphne Koller (Stanford) and Peter Norvig (Google) for agreeing to give the invited talks this year. Lastly, you try to corral 65 papers (39 Americas, 9 Asia-Pacific, and 17 Europe-Africa) into coherent groups, arrange the schedule, communicate with the chairs/committees for publications, the student sessions, workshops, demos, and local arrangements (not to mention Priscilla Rasmussen), all of whom, fortunately, are even better and more responsible than you are, and viola, here it is! You then hope that people will read this preface so that they will know what a great group put this on. Of course, no one reads these things. To test this hypothesis, the first person not connected with producing this volume who tells Eugene Charniak at the conference that he or she has read this line will get five dollars.

Oh, yes, one more thing. The program committee also picked a paper for the best paper award, a conference feature established last year. And as last year, the winner will be announced only at the last session of the conference. This implies that the winner's name(s) do not appear in the proceedings. To make sure that there is some written public record of the awards, we list below the previous year's winners.

Thanks for coming to the conference. We hope you enjoy the papers as much as we have.

Eugene Charniak
Dekang Lin
May 13, 2002

Winners of the Best Paper Award,
Conference of the Association for Computational Linguistics

2001 Ulrich Germann, Michael Jahr, Kevin Knight, Daniel Marcu, and Kenji Yamada for their paper "Fast Decoding and Optimal Decoding for Machine Translation"

Eugene Charniak for his paper "Immediate-Head Parsing for Language Models"

List of Area Chairs and Reviewers

Area Chairs

Steve Abney (AT&T Research, USA)	*Speech, Language Modeling and Statistical Methods*
Stephan Busemann (DFKI GmbH, Germany)	*Generation and Multi-Modality*
John Carroll (University of Sussex, UK)	*Parsing*
Bonnie Dorr (University of Maryland, USA)	*Lexicon and Semantics*
Jan Hajic (Charles University in Prague, Czech Republic)	*Word Segmentation, Shallow Parsing, Chunking and Tagging*
Daniel Marcu (USC/Information Sciences Institute, USA)	*Discourse and Dialogue*
Ellen Riloff (University of Utah, USA)	*NLP Applications*
Mark Steedman (University of Edinburgh, UK)	*Syntax, Grammars, Morphology and Phonology*
Keh-Yih Su (Behavior Design Corp., Taiwan)	*Translation and Multilinguality*

Reviewers

Naoki Abe (IBM T.J. Watson Research Center, USA), Anne Abeillé (Université Paris 7, France), Elisabeth André (Universität Augsburg, Germany), María Antònia Martí Antonín (Centre de Llenguatge i Computació, Spain), Susan Armstrong (ISSCO, Switzerland), Nicholas Asher (University of Texas at Austin, USA), Srinivas Bangalore (AT&T Labs-Research, USA), Leslie Barrett (Transclick Inc., USA), Roberto Basili (University of Roma, Tor Vergata, Italy), Tilman Becker (DFKI GmbH, Germany), Ibrahim Benmrad (ALAT, Tunisia), Steven Bird (Linguistic Data Consortium, University of Pennsylvania, USA), Thorsten Brants (PARC, USA), Ted Briscoe (University of Cambridge, UK), Harry Bunt (Tilburg University, The Netherlands)

Nicoletta Calzolari (Istituto di Linguistica Computazionale del CNR, Italy), Lynn Carlson (Department of Defense, USA), Bob Carpenter (SpeechWorks International, USA), David Carter (20/20 Speech, UK), Justine Cassell (MIT Media Laboratory, USA), Jing-Shin Chang (National Chi-Nan University, Taiwan), Ciprian Chelba (Microsoft Research, USA), Hsin-Hsi Chen (National Taiwan University, Taiwan), John Chen (AT&T Labs-Research, USA), Stanley F. Chen (IBM T.J. Watson Research Center, USA), Key-Sun Choi (KAIST, Korea), Jennifer Chu-Carroll (IBM, USA), Ken Church (AT&T Labs-Research, USA), Michael Collins (AT&T Labs-Research, USA), Anne Copestake (Cambridge University, UK)

Walter Daelemans (CNTS, Belgium), Ido Dagan (Bar-Ilan University, Israel), Robert Dale (Macquarie University, Australia), Tony Davis (StreamSage, Inc., USA), Eric de la Clergerie (INRIA, France), Anne De Roeck (The Open University, UK), Mona Diab (University of Maryland, College Park, USA), Barbara Di Eugenio (University of Illinois, Chicago, USA), William Dolan (Microsoft Research, USA), John Dowding (NASA Ames Research Center, USA), Mark Dras (Macquarie University, Australia), Michael Elhadad (Ben-Gurion University, Israel), Christiane Fellbaum (Princeton University, USA), Dan Flickinger (Stanford University, USA), Radu Florian (Johns Hopkins University, USA)

Rob Gaizauskas (University of Sheffield, UK), Hong Gao (Lund University, Sweden), Joshua Goodman (Microsoft Research, USA), Rebecca Green (University of Maryland, USA), Ralph Grishman (New York University, USA), Claire Grover (University of Edinburgh, UK), Nizar Habash (University of Maryland, USA), Udo Hahn (Universität Freiburg, Germany), Hilda Hardy (SUNY Albany, USA), Henk Harkema (UQAM, Canada), Marti Hearst (University of California at Berkeley, USA), Stephen Helmreich (Computing Research Laboratory/New Mexico State University, USA), Ulf Hermjakob (USC/ISI, USA), Don Hindle (Primus, USA), Julia Hirschberg (AT&T Labs-Research, USA), Graeme Hirst (University of Toronto, Canada), Eduard Hovy (USC/ISI, USA), Chu-Ren Huang (Academia Sinica, Taiwan), Rebecca Hwa (University of Maryland, USA)

Hitoshi Isahara (Communications Research Laboratory, Japan), Stephen Isard (University of Edinburgh, UK), Hongyan Jing (Lucent Technologies, Bell Labs, USA), Mark Johnson (Brown University, USA), Dominic Jones (University of Utah, USA), Aravind K. Joshi (University of Pennsylvania, USA), Laura Kallmeyer (Université Paris 7, France), Lauri Karttunen (PARC, USA), Boris Katz (MIT AI Lab, USA), Andrew Kehler (UCSD, USA), Bill Keller (University of Sussex, UK), Katrin Kirchhoff (University of Washington, USA), Judith Klavans (Columbia University, USA), Dan Klein (Stanford University, USA), Kevin Knight (USC/ISI, USA), Rob Koeling (University of Brighton, UK), Steven Krauwer (Utrecht University, The Netherlands), Hans-Ulrich Krieger (DFKI GmbH, Germany), Sadao Kurohashi (University of Tokyo, Japan), Irene Langkilde-Geary (USC/ISI, USA)

Maria Lapata (University of Edinburgh, UK), Alberto Lavelli (ITC-IRST, Italy), Alon Lavie (Carnegie Mellon University, USA), Hyo-Kyung Lee, (University of Illinois, Urbana/Champaign, USA), Lillian Lee (Cornell University, USA), Young-suk Lee (IBM Research, USA), Lori Levin (Carnegie Mellon University, USA), Elizabeth D. Liddy (Syracuse University, USA), Marc Light (MITRE, USA), Diane Litman (University of Pittsburgh, USA)

Bente Maegaard (Center for Sprogteknologi, Denmark), Christopher D. Manning (Stanford University, USA), James Martin (University of Colorado at Boulder, USA), Yuji Matsumoto (Nara Institute of Science and Technology, Japan), John Maxwell (Xerox PARC, USA), David McDonald (Mitsubishi Electric Research Labs, USA), Paola Merlo (Université de Genève, Switzerland), Adam Meyers (New York University, USA), Scott Miller (BBN Technologies, USA), Ruslan Mitkov (University of Wolverhampton, UK), Mehyrar Mohri (AT&T Labs–Research, USA), Dan Moldovan (University of Texas at Dallas, USA), Johanna Moore (University of Edinburgh, UK)

Mark-Jan Nederhof (University of Groningen, The Netherlands), Hermann Ney (Rheinisch-Westfälische Technische Hochschule Aachen, Germany), Hwee Tou Ng (National University of Singapore, Singapore), Grace Ngai (Weniwen Technologies, Hong Kong), Takashi Ninomiya (Japan Science and Technology Corporation, Japan), Stephan Oepen (YY Technologies & CSLI Stanford, USA), Karel Oliva (ÖFAI Wien, Austria), Mari Broman Olsen (Microsoft - Natural Language Group, USA), Boyan Onyshkevych (US Department of Defense, USA), Rieks op den Akker (University of Twente, The Netherlands), Miles Osborne (University of Edinburgh, UK)

Martha Palmer (University of Pennsylvania, USA), Cécile Paris (Macquarie University, Australia), Ted Pedersen (University of Minnesota, Duluth, USA), Gerald Penn (University of Toronto, Canada), Vladimir Petkevic (Charles University in Prague, Czech Republic), David Pierce (SUNY Buffalo, USA), John Prager (IBM T.J. Watson Research Center, USA), Stephen G Pulman (Oxford University, UK), Vasin Punyakanok (University of Illinois, Urbana/Champaign, USA), James Pustejovsky (Brandeis University, USA)

Owen Rambow (AT&T Labs-Research, USA), Adwait Ratnaparkhi (WhizBang! Labs, USA), Ehud Reiter (University of Aberdeen, UK), Philip Resnik (University of Maryland, USA), Christian Retoré (Université de Nantes & INRIA, France), Kiril Ribarov (Charles University in Prague, Czech Republic), Steve Richardson (Microsoft Research, USA), Thomas Rist (DFKI GmbH, Germany), Brian Roark (AT&T Labs-Research, USA), James Rogers (Earlham College, USA), Mats Rooth (Cornell University, USA), Dan Roth (University of Illinois, Urbana/Champaign, USA), Salim Roukos (IBM T.J. Watson Research Center, USA)

Kenji Sagae (Carnegie Mellon University, USA), Patrick Saint-Dizier (Institut de Recherche en Informatique de Toulouse, France), Christer Samuelsson (Inzigo Research, USA), Anoop Sarkar (Univ of Pennsylvania, USA), Giorgio Satta (Università di Padova, Italy), Hinrich Schütze (Novation Biosciences, USA), Donia Scott (University of Brighton, UK), Satoshi Sekine (New York University, USA), Vijay K. Shanker (University of Delaware, USA), Candy Sidner (Mitsubishi Electric Research Labs, USA), Sharon Small (SUNY Albany, USA), Stephen Soderland (University of Washington, USA), Harold Somers (UMIST, UK), Karen Sparck-Jones (Cambridge University, UK), Richard Sproat (AT&T Labs-Research, USA), Oliviero Stock (IRST, Italy), Michael Strube (European Media Laboratory GmbH, Germany), Tomek Strzalkowski (SUNY Albany, USA)

Scott Thomas (University of Maryland, USA), Cynthia A Thompson (University of Utah, USA), Henry S. Thompson (University of Edinburgh, UK), David Traum (USC Instititue for Creative Technologies, USA), Jun'ichi Tsujii (University of Tokyo, Japan), Carol Van Ess-Dykema (U.S. Dept. of Defense, USA), Gertjan van Noord (University of Groningen, The Netherlands), Evelyne Viegas (Microsoft Corporation, Natural Language Group, USA), Walther von Hahn (Universität Hamburg, Germany), Clare Voss (Army Research Laboratory, USA)

Wolfgang Wahlster (DFKI GmbH, Germany), Marilyn Walker (AT&T Labs-Research, USA), Nick Webb (University of Sheffield, UK), Bonnie Webber (University of Edinburgh, UK), David Weir (University of Sussex, UK), Ralph Weischedel (BBN Technologies, USA), Fuliang Weng (Robert Bosch Corp., USA), Janyce Wiebe (University of Pittsburgh, USA), Peter Wiemer-Hastings (DePaul University, USA), Yorick Wilks (University of Sheffield, UK)

David Yarowsky (Johns Hopkins University, USA), Annie Zaenen (PARC, France), Jakub Zavrel (Textkernel, The Netherlands), Daniel Zeman (Charles University in Prague, Czech Republic), Tiejun Zhao (Harbin Institute of Technology, China), Ming Zhou (Microsoft Research, Asia, China), Dav Zimak (University of Illinois, Urbana/Champaign, USA), Ingrid Zukerman (Monash University, Australia)

CONFERENCE PROGRAM

Monday, July 8

9:00 **Opening Session**

Computational Methods

9:30 *Parameter Estimation for Probabilistic Finite-State Transducers*
Jason Eisner

9:55 *Sequential Conditional Generalized Iterative Scaling*
Joshua Goodman

Generation

9:30 *Generation as Dependency Parsing*
Alexander Koller and Kristina Striegnitz

9:55 *Machine-learned Contexts for Linguistic Operations in German Sentence Realization*
Michael Gamon, Eric Ringger, Simon Corston-Oliver and Robert Moore

Question Answering

9:30 *Performance Issues and Error Analysis in an Open-Domain Question Answering System*
Dan Moldovan, Marius Pasca, Sanda Harabagiu and Mihai Surdeanu

9:55 *Learning Surface Text Patterns for a Question Answering System*
Deepak Ravichandran and Eduard Hovy

10:20 **Break**

Syntactic and Semantic Theory

10:50 *OT Syntax – Decidability of Generation-based Optimization*
Jonas Kuhn

11:15 *Phonological Comprehension and the Compilation of Optimality Theory*
Jason Eisner

11:40 *Generalized Encoding of Description Spaces and its Application to Typed Feature Structures*
Gerald Penn

12:05 *Ellipsis Resolution with Underspecified Scope*
Michael Schiehlen

Coreference Resolution and Generation

10:50 *Resolving Pronominal Reference to Abstract Entities*
Donna K. Byron

11:15 *Pronominalization in Generated Discourse and Dialogue*
Charles B. Callaway and James L. Lester

11:40 *Generating Minimal Definite Descriptions*
Claire Gardent

12:05 *Improving Machine Learning Approaches to Coreference Resolution*
Vincent Ng and Claire Cardie

Specialized Parsing and Grammar Induction

10:50 *Parsing Non-recursive CFGs*
Mark-Jan Nederhof and Giorgio Satta

11:15 *Active Learning for Statistical Natural Language Parsing*
Min Tang, Xiaoqiang Luo and Salim Roukos

11:40 *A Generative Constituent-Context Model for Improved Grammar Induction*
Dan Klein and Christopher D. Manning

12:05 *A Simple Pattern-matching Algorithm for Recovering Empty Nodes and Their Antecedents*
Mark Johnson

12:30 **Lunch**

Applications

14:00 *Pronunciation Modeling for Improved Spelling Correction*
Kristina Toutanova and Robert Moore

14:25 *Measuring Text Reuse*
Paul Clough, Robert Gaizauskas, Scott S.L. Piao and Yorick Wilks

14:50 *Semi-Supervised Maximum Entropy Based Approach to Acronym and Abbreviation Normalization in Medical Texts*
Serguei Pakhomov

15:15 *A Framework and Graphical Development Environment for Robust NLP Tools and Applications*
Hamish Cunningham, Diana Maynard, Kalina Bontcheva and Valentin Tablan

Language Modeling

14:00 *Improving Language Model Size Reduction Using Better Pruning Criteria*
Jianfeng Gao and Min Zhang

14:25 *Exploring Asymmetric Clustering for Statistical Language Modeling*
Jianfeng Gao, Joshua Goodman, Guihong Cao and Hang Li

14:50 *A Study on Richer Syntactic Dependencies for Structured Language Modeling*
Peng Xu, Ciprian Chelba and Frederick Jelinek

15:15 *Entropy Rate Constancy in Text*
Dmitriy Genzel and Eugene Charniak

Verb Meaning

14:00 *A Multilingual Paradigm for Automatic Verb Classification*
Paola Merlo, Suzanne Stevenson, Vivian Tsang and Gianluca Allaria

14:25 *Verb Paraphrase Based on Case Frame Alignment*
Nobuhiro Kaji, Daisuke Kawahara, Sadao Kurohashi and Satoshi Sato

14:50 *Inducing German Semantic Verb Classes from Purely Syntactic Subcategorisation Information*
Sabine Schulte im Walde and Chris Brew

15:40 **Break**

16:10 **The ACL: 40 Years of Publication**

16:25: **Lifetime Achievement Award**

Tuesday, July 9

Relations and Meaning

9:00 *Scaling Context Space*
James Curran and Marc Moens

9:25 *The Necessity of Parsing for Predicate Argument Recognition*
Daniel Gildea and Martha Palmer

9:50 *The Descent of Hierarchy, and Selection in Relational Semantics*
Barbara Rosario, Marti Hearst and Charles Fillmore

10:15 *An Unsupervised Method for Word Sense Tagging Using Parallel Corpora*
Mona Diab and Philip Resnik

Beyond Standard CFG Parsing

9:00 *New Ranking Algorithms for Parsing and Tagging: Kernels Over Discrete Structures, and the Voted Perceptron*
Michael Collins and Nigel Duffy

9:25 *Parsing the Wall Street Journal Using a Lexical-Functional Grammar and Discriminative Estimation Techniques*
Stefan Riezler, Tracy H. King, Ronald M. Kaplan, Richard Crouch, John T. Maxwell III and Mark Johnson

9:50 *Dynamic Programming for Parsing and Estimation of Stochastic Unification-based Grammars*
Stuart Geman and Mark Johnson

10:15 *Markov Parsing: Lattice Rescoring with a Statistical Parser*
Brian Roark

9:00 Demo Session I

9:50 Demo Session II

10:40 **Break**

11:20 **ACL Business Meeting**

12:20 **Lunch**

Machine Translation and Evaluation

14:00 *Discriminative Training and Maximum Entropy Models for Statistical Machine Translation*
Franz Josef Och and Hermann Ney

14:25 *A Decoder for Syntax-based Statistical MT*
Kenji Yamada and Kevin Knight

14:50 *Bleu: a Method for Automatic Evaluation of Machine Translation*
Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu

Combinatory Categorial Grammar

14:00 *Coupling CCG and Hybrid Logic Dependency Semantics*
Jason Baldridge and Geert-Jan Kruijff

14:25 *Building Deep Dependency Structures Using a Wide-Coverage CCG Parser*
Stephen Clark, Julia Hockenmaier and Mark Steedman

14:50 *Generative Models for Statistical Parsing with Combinatory Categorial Grammar*
Julia Hockenmaier and Mark Steedman

14:00 Demo Session III

15:15 **Break**

15:45 **Invited Talk:** *Probabilistic Models of Relational Data*
Daphne Koller

Bootstrapping

16:55 *Word Translation Disambiguation Using Bilingual Bootstrapping*
Cong Li and Hang Li

17:20 *Applying Co-training to Reference Resolution*
Christoph Mueller, Stefan Rapp and Michael Strube

17:45 *Bootstrapping*
Steven Abney

Discourse and Dialogue

16:55 *An Unsupervised Approach to Recognizing Discourse Relations*
Daniel Marcu and Abdessamad Echihabi

17:20 *MATCH: An Architecture for Multimodal Dialogue Systems*
Michael Johnston, Srinivas Bangalore, Gunaranjan Vasireddy, Amanda Stent, Patrick Ehlen, Marilyn Walker, Steve Whittaker and Preetam Maloor

17:45 *What's the Problem: Automatically Identifying Problematic Dialogues in DARPA Communicator Dialogue Systems*
Helen Wright Hastie, Rashmi Prasad and Marilyn Walker

Machine Translation Building Blocks

16:55 *Evaluating Translational Correspondence Using Annotation Projection*
Rebecca Hwa, Philip Resnik, Amy Weinberg and Okan Kolak

17:20 *Translating Named Entities Using Monolingual and Bilingual Resources*
Yaser Al-Onaizan and Kevin Knight

17:45 *Using Similarity Scoring to Improve the Bilingual Dictionary for Sub-sentential Alignment*
Katharina Probst and Ralf Brown

Wednesday, July 10

9:00 **Student Research Workshop**

11:00 **Break**

Using the Web

11:30 *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*
Peter Turney

11:55 *Is It the Right Answer? Exploiting Web Redundancy for Answer Validation*
Bernardo Magnini, Matteo Negri, Roberto Prevete and Hristo Tanev

Shallow (and Deep) Parsing

11:30 *Shallow Parsing on the Basis of Words Only: A Case Study*
Antal van den Bosch and Sabine Buchholz

11:55 *An Integrated Archictecture for Shallow and Deep Processing*
Berthold Crysmann, Anette Frank, Kiefer Bernd, Stefan Mueller, Guenter Neumann, Jakub Piskorski, Ulrich SCchaefer, Melanie Siegel, Hans Uszkoreit, Feiyu Xu, Markus Becker and Hans-Ulrich Krieger

11:30 Demo Session IV

12:20 **Lunch**

Compression and Summarization

14:00 *A Noisy-Channel Model for Document Compression*
Hal Daume III and Daniel Marcu

14:25 *From Single to Multi-document Summarization*
Chin-Yew Lin and Eduard Hovy

14:50 *Supervised Ranking in Open-Domain Text Summarization*
Tadashi Nomoto and Yuji Matsumoto

Named Entity Recognition

14:00 *Named Entity Recognition Using an HMM-based Chunk Tagger*
GuoDong ZHOU and Jian SU

14:25 *Teaching a Weaker Classifier: Named Entity Recognition on Upper Case Text*
Hai Leong Chieu and Hwee Tou Ng

14:50 *Ranking Algorithms for Named Entity Extraction: Boosting and the Voted Perceptron*
Michael Collins

Words and Below

14:00 *Revision Learning and its Application to Part-of-Speech Tagging*
Tetsuji Nakagawa, Taku Kudo and Yuji Matsumoto

14:25 *An Empirical Study of Active Learning with Support Vector Machines for Japanese Word Segmentation*
Manabu Sassano

14:50 *Memory-Based Learning of Morphology with Stochastic Transducers*
Alexander Clark

15:15 **Break**

15:45 **Invited Talk:** *Better Web Search, with and without Computational Linguistics*
Peter Norvig

16:50 **Best Paper Award and Closing**

Table of Contents

Parameter Estimation for Probabilistic Finite-State Transducers
Jason Eisner 1

Sequential Conditional Generalized Iterative Scaling
Joshua Goodman 9

Generation as Dependency Parsing
Alexander Koller and Kristina Striegnitz 17

Machine-learned Contexts for Linguistic Operations in German Sentence Realization
Michael Gamon, Eric Ringger, Simon Corston-Oliver and Robert Moore 25

Performance Issues and Error Analysis in an Open-Domain Question Answering System
Dan Moldovan, Marius Pasca, Sanda Harabagiu and Mihai Surdeanu 33

Learning Surface Text Patterns for a Question Answering System
Deepak Ravichandran and Eduard Hovy 41

OT Syntax – Decidability of Generation-based Optimization
Jonas Kuhn 48

Phonological Comprehension and the Compilation of Optimality Theory
Jason Eisner 56

Generalized Encoding of Description Spaces and its Application to Typed Feature Structures
Gerald Penn 64

Ellipsis Resolution with Underspecified Scope
Michael Schiehlen 72

Resolving Pronominal Reference to Abstract Entities
Donna K. Byron 80

Pronominalization in Generated Discourse and Dialogue
Charles B. Callaway and James L. Lester 88

Generating Minimal Definite Descriptions
Claire Gardent 96

Improving Machine Learning Approaches to Coreference Resolution
Vincent Ng and Claire Cardie 104

Parsing Non-recursive CFGs
Mark-Jan Nederhof and Giorgio Satta 112

Active Learning for Statistical Natural Language Parsing
Min Tang, Xiaoqiang Luo and Salim Roukos 120

A Generative Constituent-Context Model for Improved Grammar Induction
Dan Klein and Christopher D. Manning 128

A Simple Pattern-matching Algorithm for Recovering Empty Nodes and their Antecedents
Mark Johnson 136

Pronunciation Modeling for Improved Spelling Correction
Kristina Toutanova and Robert Moore 144

Measuring Text Reuse
Paul Clough, Robert Gaizauskas, Scott S.L. Piao and Yorick Wilks 152

Semi-Supervised Maximum Entropy Based Approach to Acronym and Abbreviation Normalization in Medical Texts
Serguei Pakhomov 160

A Framework and Graphical Development Environment for Robust NLP Tools and Applications
Hamish Cunningham, Diana Maynard, Kalina Bontcheva and Valentin Tablan 168

Improving Language Model Size Reduction Using Better Pruning Criteria
Jianfeng Gao and Min Zhang 176

Exploring Asymmetric Clustering for Statistical Language Modeling
Jianfeng Gao, Joshua Goodman, Guihong Cao and Hang Li 183

A Study on Richer Syntactic Dependencies for Structured Language Modeling
Peng Xu, Ciprian Chelba and Frederick Jelinek 191

Entropy Rate Constancy in Text
Dmitriy Genzel and Eugene Charniak 199

A Multilingual Paradigm for Automatic Verb Classification
Paola Merlo, Suzanne Stevenson, Vivian Tsang and Gianluca Allaria 207

Verb Paraphrase Based on Case Frame Alignment
Nobuhiro Kaji, Daisuke Kawahara, Sadao Kurohashi and Satoshi Sato 215

Inducing German Semantic Verb Classes from Purely Syntactic Subcategorisation Information
Sabine Schulte im Walde and Chris Brew 223

Scaling Context Space
James Curran and Marc Moens 231

The Necessity of Parsing for Predicate Argument Recognition
Daniel Gildea and Martha Palmer 239

The Descent of Hierarchy, and Selection in Relational Semantics
Barbara Rosario, Marti Hearst and Charles Fillmore 247

An Unsupervised Method for Word Sense Tagging Using Parallel Corpora
Mona Diab and Philip Resnik 255

New Ranking Algorithms for Parsing and Tagging: Kernels over Discrete Structures, and the Voted Perceptron
Michael Collins and Nigel Duffy 263

Parsing the Wall Street Journal Using a Lexical-Functional Grammar and Discriminative Estimation Techniques
Stefan Riezler, Tracy H. King, Ronald M. Kaplan, Richard Crouch, John T. Maxwell III and Mark Johnson 271

Dynamic Programming for Parsing and Estimation of Stochastic Unification-based Grammars
Stuart Geman and Mark Johnson 279

Markov Parsing: Lattice Rescoring with a Statistical Parser
Brian Roark 287

Discriminative Training and Maximum Entropy Models for Statistical Machine Translation
Franz Josef Och and Hermann Ney 295

A Decoder for Syntax-based Statistical MT
Kenji Yamada and Kevin Knight 303

Bleu: a Method for Automatic Evaluation of Machine Translation
Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu 311

Coupling CCG and Hybrid Logic Dependency Semantics
Jason Baldridge and Geert-Jan Kruijff 319

Building Deep Dependency Structures Using a Wide-Coverage CCG Parser
Stephen Clark, Julia Hockenmaier and Mark Steedman 327

Generative Models for Statistical Parsing with Combinatory Categorial Grammar
Julia Hockenmaier and Mark Steedman 335

Word Translation Disambiguation Using Bilingual Bootstrapping
Cong Li and Hang Li 343

Applying Co-Training to Reference Resolution
Christoph Mueller, Stefan Rapp and Michael Strube 352

Bootstrapping
Steven Abney 360

An Unsupervised Approach to Recognizing Discourse Relations
Daniel Marcu and Abdessamad Echihabi 368

MATCH: An Architecture for Multimodal Dialogue Systems
Michael Johnston, Srinivas Bangalore, Gunaranjan Vasireddy, Amanda Stent, Patrick Ehlen, Marilyn Walker, Steve Whittaker and Preetam Maloor 376

What's the Problem: Automatically Identifying Problematic Dialogues in DARPA Communicator Dialogue Systems
Helen Wright Hastie, Rashmi Prasad and Marilyn Walker 384

Evaluating Translational Correspondence Using Annotation Projection
Rebecca Hwa, Philip Resnik, Amy Weinberg and Okan Kolak 392

Translating Named Entities Using Monolingual and Bilingual Resources
Yaser Al-Onaizan and Kevin Knight 400

Using Similarity Scoring to Improve the Bilingual Dictionary for Sub-sentential Alignment
Katharina Probst and Ralf Brown 409

Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews
Peter Turney 417

Is It the Right Answer? Exploiting Web Redundancy for Answer Validation
Bernardo Magnini, Matteo Negri, Roberto Prevete and Hristo Tanev 425

Shallow Parsing on the Basis of Words Only: A Case Study
Antal van den Bosch and Sabine Buchholz 433

An Integrated Archictecture for Shallow and Deep Processing
Berthold Crysmann, Anette Frank, Kiefer Bernd, Stefan Mueller, Guenter Neumann, Jakub Piskorski, Ulrich SCchaefer, Melanie Siegel, Hans Uszkoreit, Feiyu Xu, Markus Becker and Hans-Ulrich Krieger 441

A Noisy-Channel Model for Document Compression
Hal Daume III and Daniel Marcu 449

From Single to Multi-document Summarization
Chin-Yew Lin and Eduard Hovy 457

Supervised Ranking in Open-Domain Text Summarization
Tadashi Nomoto and Yuji Matsumoto 465

Named Entity Recognition Using an HMM-based Chunk Tagger
GuoDong ZHOU and Jian SU 473

Teaching a Weaker Classifier: Named Entity Recognition on Upper Case Text
Hai Leong Chieu and Hwee Tou Ng 481

Ranking Algorithms for Named Entity Extraction: Boosting and the Voted Perceptron
Michael Collins 489

Revision Learning and its Application to Part-of-Speech Tagging
Tetsuji Nakagawa, Taku Kudo and Yuji Matsumoto 497

An Empirical Study of Active Learning with Support Vector Machines for Japanese Word Segmentation
Manabu Sassano 505

Memory-Based Learning of Morphology with Stochastic Transducers
Alexander Clark 513

Author Index

Abney, Steven 360
Allaria, Gianluca 207
Al-Onaizan, Yaser 400
Baldridge, Jason 319
Bangalore, Srinivas 376
Becker, Markus 441
Bernd, Kiefer 441
Bontcheva, Kalina 168
Brew, Chris 223
Brown, Ralf 409
Buchholz, Sabine 433
Byron, Donna K. 80
Callaway, Charles B. 88
Cao, Guihong 183
Cardie, Claire 104
Charniak, Eugene 199
Chelba, Ciprian 191
Chieu, Hai Leong 481
Clark, Alexander 513
Clark, Stephen 327
Clough, Paul 152
Collins, Michael 263, 489
Corston-Oliver, Simon 25
Crouch, Richard 271
Crysmann, Berthold 441
Cunningham, Hamish 168
Curran, James 231
Daume III, Hal 449
Diab, Mona 255
Duffy, Nigel 263
Echihabi, Abdessamad 368
Ehlen, Patrick 376
Eisner, Jason 1, 56
Fillmore, Charles 247
Frank, Anette 441
Gaizauskas, Robert 152
Gamon, Michael 25
Gao, Jianfeng 176, 183
Gardent, Claire 96
Geman, Stuart 279
Genzel, Dmitriy 199
Gildea, Daniel 239
Goodman, Joshua 9, 183
Harabagiu, Sanda 33
Hastie, Helen Wright 384
Hearst, Marti 247
Hockenmaier, Julia 327, 335
Hovy, Eduard 41, 457
Hwa, Rebecca 392
Jelinek, Frederick 191
Johnson, Mark 136, 271, 279
Johnston, Michael 376
Kaji, Nobuhiro 215
Kaplan, Ronald M. 271
Kawahara, Daisuke 215
King, Tracy H. 271
Klein, Dan 128
Knight, Kevin 303, 400
Kolak, Okan 392
Koller, Alexander 17
Krieger, Hans-Ulrich 441
Kruijff, Geert-Jan 319
Kudo, Taku 497
Kuhn, Jonas 48
Kurohashi, Sadao 215
Lester, James L. 88
Li, Cong 343
Li, Hang 183, 343
Lin, Chin-Yew 457
Luo, Xiaoqiang 120
Magnini, Bernardo 425
Maloor, Preetam 376
Manning, Christopher D. 128
Marcu, Daniel 368, 449
Matsumoto, Yuji 465, 497
Maxwell, John T. III 271
Maynard, Diana 168
Merlo, Paola 207
Moens, Marc 231
Moldovan, Dan 33
Moore, Robert 25, 144
Mueller, Christoph 352
Mueller, Stefan 441
Nakagawa, Tetsuji 497
Nederhof, Mark-Jan 112
Negri, Matteo 425
Neumann, Guenter 441
Ney, Hermann 295
Ng, Hwee Tou 481
Ng, Vincent 104
Nomoto, Tadashi 465
Och, Franz Josef 295

Pakhomov, Serguei 160
Palmer, Martha 239
Papineni, Kishore 311
Pasca, Marius 33
Penn, Gerald 64
Piao, Scott S.L. 152
Piskorski, Jakub 441
Prasad, Rashmi 384
Prevete, Roberto 425
Probst, Katharina 409
Rapp, Stefan 352
Ravichandran, Deepak 41
Resnik, Philip 255, 392
Riezler, Stefan 271
Ringger, Eric 25
Roark, Brian 287
Rosario, Barbara 247
Roukos, Salim 120
Roukos, Salim 311
Sassano, Manabu 505
Sato, Satoshi 215
Satta, Giorgio 112
SCchaefer, Ulrich 441
Schiehlen, Michael 72
Schulte im Walde, Sabine 223
Siegel, Melanie 441
Steedman, Mark 327, 335
Stent, Amanda 376
Stevenson, Suzanne 207
Striegnitz, Kristina 17
Strube, Michael 352
SU, Jian 473
Surdeanu, Mihai 33
Tablan, Valentin 168
Tanev, Hristo 425
Tang, Min 120
Toutanova, Kristina 144
Tsang, Vivian 207
Turney, Peter 417
Uszkoreit, Hans 441
van den Bosch, Antal 433
Vasireddy, Gunaranjan 376
Walker, Marilyn 376
Walker, Marilyn 384
Ward, Todd 311
Weinberg, Amy 392
Whittaker, Steve 376
Wilks, Yorick 152
Xu, Feiyu 441
Xu, Peng 191
Yamada, Kenji 303
Zhang, Min 176
ZHOU, GuoDong 473
Zhu, Wei-Jing 311