

Association for Computational Linguistics

1st Meeting of the North American Chapter of the Association for Computational Linguistics

Proceedings of the Conference

April 29—May 4, 2000
Seattle, Washington, USA

NAACL 2000—PREFACE

On behalf of the Program Committee for NAACL 2000, I am pleased to present you with the papers accepted for presentation at the First Meeting of the North American Chapter of the Association for Computational Linguistics, held in Seattle, Washington, April 29-May 4, 2000.

NAACL received a gratifyingly large number of papers from around the world. Submissions were received from 28 countries. Reviewing was blind to all reviewers and area chairs. It was also highly selective. Out of 166 submissions, 43 were selected for presentation at NAACL 2000.

Selecting the papers was not an easy task. In total, over 110 reviewers, representing 20 different countries, reported to a senior program committee consisting of eight area chairs. The senior program committee spent an intensive day at a meeting in Virginia reaching the final decisions. The area chairs and reviewers cannot be thanked enough for the conscientious and painstaking jobs they performed. All those who contributed are named on the following page, but I would particularly like to express my thanks here to the area chairs: Michael Collins (AT&T Labs - Research), Nancy Green (University of North Carolina at Greensboro), Graeme Hirst (University of Toronto), Kevin Knight (USC/Information Sciences Institute), Dekang Lin (University of Manitoba), Diane Litman (AT&T Labs - Research), Philip Resnik (University of Maryland), and Andreas Stolcke (SRI International).

In its first year, it was advantageous to co-locate NAACL with ANLP, an established conference. To coordinate the two conferences, submissions focusing on end-applications were invited to ANLP 2000, while submissions focusing on methodology were invited to NAACL 2000. Future NAACL conferences will encourage both types of submissions.

Many thanks to Marie Meteer for her outstanding service and support in her role as General Chair.

We are extremely grateful to Linda Fresques for her work on all aspects of the process, from managing the submissions to preparing the proceedings, and to Eli Kane for developing and maintaining the review assignment and submission software. We also thank Lori Allen, Jacob Feddersen, and Nayana Datta for help with the submission process. In addition, we thank the Computing Research Laboratory at New Mexico State University for providing significant internal administrative and technical support.

Finally, we would like to express our gratitude to Kathy McCoy and Priscilla Rasmussen for their consistent support and guidance, and to the ACL and NAACL Executive Committees.

Welcome to NAACL 2000!

Janyce Wiebe
Program Chair
April 2000

NAACL Program Committee

Program Committee Chair:

Janyce Wiebe, New Mexico State University

Area Chairs:

Michael Collins, AT&T Labs - Research
Nancy Green, University of North Carolina at Greensboro
Graeme Hirst, University of Toronto
Kevin Knight, USC/Information Sciences Institute
Dekang Lin, University of Manitoba
Diane Litman, AT&T Labs - Research
Philip Resnik, University of Maryland
Andreas Stolcke, SRI International

Reviewers:

Steve Abney, AT&T Labs-Research
Chinatsu Aone, SRA Corporation
Steven Bird, University of Pennsylvania
Francis Bond, NTT
Eric Brill, Microsoft
Ted Briscoe, University of Cambridge
Rebecca Bruce, University of North Carolina at Asheville
Sandra Carberry, University of Delaware
Claire Cardie, Cornell University
Jean Carletta, University of Edinburgh
Bob Carpenter, SpeechWorks International
John Carroll, University of Sussex
Patrick Caudal, Université Paris
Justine Cassell, Massachusetts Institute of Technology
Eugene Charniak, Brown University
Ciprian Chelba, Johns Hopkins University
Jennifer Chu-Carroll, Bell Labs, Lucent Technologies
Kenneth W. Church, AT&T Labs - Research
Phil Cohen, Oregon Graduate Institute of Science and Technology
Walter Daelemans, University of Antwerp
Ido Dagan, Bar-Ilan University
Mary Dalrymple, Xerox PARC
Barbara DiEugenio, University of Illinois at Chicago
Bonnie J. Dorr, University of Maryland
Michael Elhadad, Ben-Gurion University of the Negev
Thierry Fontenelle, Luxembourg
Robert Gaizauskas, University of Sheffield
Claire Gardent, University of the Saarland

Oren Glickman, Bar-Ilan University
Joshua Goodman, Microsoft Research
Gregory Grefenstette, Xerox Research Centre Europe
Udo Hahn, University of Freiburg
Jan Hajič, Charles University
Eva Hajičova, Charles University
Christine Halverson, SRI International
Mary Harper, Purdue University
Peter Heeman, Oregon Graduate Institute of Science and Technology
Jamie Henderson, University of Exeter
Ulf Hermjakob, USC/Information Sciences Institute
Vasileios Hatzivassiloglou, Columbia University
Julia Hirschberg, AT&T Labs - Research
Beth-Ann Hockey, NASA/RIACS
Rukmini Iyer, BBN Technologies
Mark Johnson, Brown University
Doug Jones, Department of Defense
Arne Jonsson, Linköping University
Sanjeev Khudanpur, Johns Hopkins University
Adam Kilgarriff, University of Brighton
George Kiraz, Bell Labs, Lucent Technologies
Judith Klavans, Columbia University
Alistair Knott, University of Otago
Yuval Krymolowski, Bar-Ilan University
Sadao Kurohashi, Kyoto University
John Lafferty, Carnegie Mellon University
Alon Lavie, Carnegie Mellon University
Claudia Leacock, Educational Testing Service
Lilian Lee, Cornell University
James Lester, North Carolina State University
Lori Levin, Carnegie Mellon University
Hang Li, NEC

Ken Litkowski, CL Research
Mike Macon, Oregon Graduate Institute of
Science and Technology
Inderjeet Mani, Mitre Corporation
Chris Manning, Stanford University
Daniel Marcu, USC/Information Sciences
Institute
Zvi Marx, Bar-Ilan University
Yuji Matsumoto, Nara Institute of Science and
Technology
Susan McRoy, University of Wisconsin-
Milwaukee
Dan Melamed, West Group
Paola Merlo, University of Geneva
Vibhu Mittal, JPRC
Mehryar Mohri, AT&T Labs-Research
Johanna Moore, University of Edinburgh
Hwee Tou Ng, DSO National Laboratories,
Singapore
Christine Nakatani, Bell Labs, Lucent
Technologies
Doug Oard, University of Maryland
Martha Palmer, University of Pennsylvania
Cecile Paris, CSIRO Australia
Dragomir Radev, University of Michigan
Adwait Rathnaparkhi, IBM
Manny Rayner, NASA/RIACS
Jeff Reynar, Microsoft
Klaus Ries, Carnegie Mellon University
Ellen Riloff, University of Utah
Dan Roth, University of Illinois at Urbana-
Champaign
Anoop Saakar, University of Pennsylvania
Mark Sanderson, University of Sheffield
Antonio Sanfilippo, European Commission, DG
Information Society
Giorgio Satta, University of Padua
Stuart Shieber, Harvard University
Lenhart Schubert, University of Rochester
Donia Scott, Information Technology Research
Institute
Satoshi Sekine, New York University
Kristie Seymore, Carnegie Mellon University
Liz Shriberg, SRI International
Karen Sparck-Jones, Cambridge University
Srinivas Bangalore, AT&T Labs-Research
Manfred Stede, Technische Universitat Berlin
Mark Steedman, University of Edinburgh
Suzanne Stevenson, Rutgers University
Richard Sproat, AT&T Labs-Research
Matthew Stone, Rutgers University

Michael Strube, University of Edinburgh
Keh-Yih Su, Behavior Design Corporation
Marc Swerts, IPO Eindhoven
Paul Taylor, University of Edinburgh
David Traum, University of Maryland
Lucy Vanderwende, Microsoft Research
Piek Vossen, Sail Labs
Bonnie Webber, University of Edinburgh
David Weir, University of Sussex
Steve Whittaker, AT&T Labs - Research
David Yarowsky, Johns Hopkins University

NAACL Table of Contents

Section 2: North American Chapter of the Association for Computational Linguistics (NAACL)

NAACL Preface and List of Reviewers

Janyce Wiebe, Program Committee Chair NAACL*i*

Modelling Grounding and Discourse Obligations Using Update Rules

Colin Matheson, Massimo Poesio, David Traum 1

The Automatic Translation of Discourse Structures

Daniel Marcu, Lynn Carlson, Maki Watanabe 9

A Probabilistic Genre-Independent Model of Pronominalization

Michael Strube, Maria Wolters 18

Advances in Domain Independent Linear Text Segmentation

Freddy Y.Y. Choi 26

Bagging and Boosting a Treebank Parser

John C. Henderson, Eric Brill 34

Encoding Information on Adjectives in a Lexical-Semantic Net for Computational Applications

Antonietta Alonge, Francesca Bertagna, Nicoletta Calzolari, Adriana Roventini, Antonio Zampolli 42

Noun Phrase Recognition by System Combination

Erik F. Tjong Kim Sang 50

The FrameNet Tagset for Frame-Semantic and Syntactic Coding of Predicate-Argument Structure

Christopher Johnson, Charles J. Fillmore 56

A Simple Approach to Building Ensembles of Naïve Bayesian Classifiers for Word Sense Disambiguation

Ted Pedersen 63

Generation in the Lambek Calculus Framework: An Approach with Semantic Proof Nets

Sylvain Pogodalla 70

Word-for-Word Glossing with Contextually Similar Words

Patrick Pantel, Dekang Lin 78

Arabic Morphology Generation Using a Concatenative Strategy

Violetta Cavalli-Sforza, Abdelhadi Soudi, Teruko Mitamura 86

Morphological Tagging: Data vs. Dictionaries

Jan Hajič 94

<i>The Effectiveness of Corpus-Induced Dependency Grammars for Post-Processing Speech</i>	
M.P. Harper, C.M. White, W. Wang, M.T. Johnson, R.A. Helzerman	102
<i>Analyzing Dependencies of Japanese Subordinate Clauses Based on Statistics of Scope Embedding Preference</i>	
Takehito Utsuro, Shigeyuki Nishiokayama, Masakazu Fujio, Yuji Matsumoto	110
<i>Rapid Parser Development: A Machine Learning Approach for Korean</i>	
Ulf Hermjakob	118
<i>A Classification Approach to Word Prediction</i>	
Yair Even-Zohar, Dan Roth	124
<i>A Maximum-Entropy-Inspired Parser</i>	
Eugene Charniak	132
<i>An Unsupervised Method for Detecting Grammatical Errors</i>	
Martin Chodorow, Claudia Leacock	140
<i>Detecting Errors within a Corpus Using Anomaly Detection</i>	
Eleazar Eskin	148
<i>Exploiting Auxiliary Distributions in Stochastic Unification-Based Grammars</i>	
Mark Johnson, Stefan Riezler	154
<i>Ambiguity Packing in Constraint-Based Parsing—Practical Results</i>	
Stephan Oepen, John Carroll	162
<i>Forest-Based Statistical Sentence Generation</i>	
Irene Langkilde	170
<i>Cut and Paste Based Text Summarization</i>	
Hongyan Jing, Kathleen R. McKeown	178
<i>Minimizing Word Error Rate in Textual Summaries of Spoken Language</i>	
Klaus Zechner, Alex Waibel	186
<i>Trainable Methods for Surface Natural Language Generation</i>	
Adwait Ratnaparkhi	194
<i>Evaluating Automatic Dialogue Strategy Adaptation for a Spoken Dialogue System</i>	
Jennifer Chu-Carroll, Jill Suzanne Nickerson	202
<i>Learning to Predict Problematic Situations in a Spoken Dialogue System: Experiments with How May I Help You?</i>	
Marilyn Walker, Irene Langkilde, Jerry Wright, Allen Gorin, Diane Litman	210

<i>Predicting Automatic Speech Recognition Performance Using Prosodic Cues</i> Diane J. Litman, Julia B. Hirschberg, Marc Swerts	218
<i>A Novel Use of Statistical Parsing to Extract Information from Text</i> Scott Miller, Heidi Fox, Lance Ramshaw, Ralph Weischedel	226
<i>Assigning Function Tags to Parsed Text</i> Don Blaheta, Eugene Charniak	234
<i>Mostly-Unsupervised Statistical Segmentation of Japanese: Applications to Kanji</i> Rie Kubota Ando, Lillian Lee	241
<i>Removing Left Recursion from Context-Free Grammars</i> Robert C. Moore	249
<i>Using Semantic Preferences to Identify Verbal Participation in Role Switching Alternations</i> Diana McCarthy	256
<i>Tagging Sentence Boundaries</i> Andrei Mikheev	264
<i>Left-to-Right Parsing and Bilexical Context-Free Grammars</i> Mark-Jan Nederhof, Giorgio Satta	272
<i>Acknowledgements in Human-Computer Interaction</i> Karen Ward, Peter A. Heeman	280
<i>A New Algorithm for the Alignment of Phonetic Sequences</i> Grzegorz Kondrak	288
<i>Finite-State Reduplication in One-Level Prosodic Morphology</i> Markus Walther	296
<i>A Finite State and Data-Oriented Method for Grapheme to Phoneme Conversion</i> Gosse Bouma	303
<i>A Framework for Robust Semantic Interpretation</i> Carolyn P. Rosé	311
<i>Understanding “Each Other”</i> Claire Gardent, Karsten Konrad	319
<i>An Empirical Assessment of Semantic Interpretation</i> Martin Romacker, Udo Hahn	327

