

**Proceedings of the
Second Conference
on
Empirical Methods in
Natural Language Processing**

**Sponsored by
The Association for Computational Linguistics
and
SIGDAT, a Special Interest Group of the ACL**

Held in cooperation with AAI-97

**Edited by
Claire Cardie
and
Ralph Weischedel**

**August 1-2, 1997
Brown University
Providence, Rhode Island, USA**

Proceedings of the
Second Conference
on
Empirical Methods in
Natural Language Processing

Sponsored by
The Association for Computational Linguistics
and
SIGDAT, a Special Interest Group of the ACL

Held in cooperation with AAI-97

Edited by
Claire Cardie
and
Ralph Weischedel

August 1-2, 1997
Brown University
Providence, Rhode Island, USA

Order additional copies from:

ACL
P.O. Box 6090
Somerset, NJ 08875
1-908-873-3898
acl@bellcore.com

SPONSORS:

The Association for Computational Linguistics (ACL)
SIGDAT (ACL's SIG for Linguistic Data and Corpus-based Approaches to NLP)

INVITED SPEAKERS:

Tom Mitchell
Michael Mauldin

ORGANIZERS:

Claire Cardie, Chair
Ralph Weischedel, Co-chair

LOCAL ARRANGEMENTS CHAIR:

Eugene Charniak

PROGRAM COMMITTEE:

Ted Briscoe	(Cambridge University)
Rebecca Bruce	(Southern Methodist University)
Michael Collins	(University of Pennsylvania)
Bruce Croft	(University of Massachusetts, Amherst)
Carl de Marcken	(Massachusetts Institute of Technology)
Joshua Goodman	(Harvard University)
Eduard Hovy	(USC Information Sciences Institute)
Nancy Ide	(Vassar College)
K.L. Kwok	(Queens College, CUNY)
John Lafferty	(Carnegie Mellon University)
Steve Maiorano	(ORD)
Kemal Oflazer	(Bilkent University)
Philip Resnik	(University of Maryland)
Dekai Wu	(Hong Kong University of Science and Technology)
David Yarowsky	(Johns Hopkins University)

FURTHER INFORMATION:

Claire Cardie
Department of Computer Science
Cornell University
4142 Upson Hall
Ithaca, NY 14850 USA
e-mail: cardie@cs.cornell.edu

Ralph Weischedel
BBN Systems and Technologies
70 Fawcett Street
Cambridge, MA 02138
USA
e-mail: weischedel@bbn.com

CONFERENCE PROGRAM

Friday, August 1

- 8:00 - 9:00 Registration and Continental Breakfast
- 9:00 - 9:15 Welcome
- 9:15 - 9:40 Adwait Ratnaparkhi
A Linear Observed Time Statistical Parser Based on Maximum Entropy Models
- 9:40 - 10:05 Joshua Goodman
Global Thresholding and Multiple-Pass Parsing
- 10:05 - 10:30 Carolyn Penstein Rosé and Alon Lavie
An Efficient Distribution of Labor in a Two Stage Robust Interpretation Process
- 10:30 - 11:00 Break (held jointly with the Uncertainty in AI Conference)
- 11:00 - 11:25 Doug Beeferman, Adam Berger, and John Lafferty
Text Segmentation Using Exponential Models
- 11:25 - 11:50 Korin Richmond, Andrew Smith, and Einat Amitay
Detecting Subject Boundaries Within Text: A Language Independent Statistical Approach
- 11:50 - 12:15 Ido Dagan, Yael Karov, and Dan Roth
Mistake-Driven Learning in Text Categorization
- 12:15 - 1:45 CATERED LUNCH
- 1:45 - 2:45 INVITED TALK, Tom Mitchell
Machine Learning and Extracting Information from the Web
- 2:45 - 3:00 Break
- 3:00 - 3:25 Thorsten Brants, Wojciech Skut, and Brigitte Krenn
Tagging Grammatical Functions
- 3:25 - 3:50 Jo Calder
On Aligning Trees
- 3:50 - 4:15 Break
- 4:15 - 4:40 Lawrence Saul and Fernando Pereira
Aggregate and Mixed-order Markov Models for Statistical Language Processing
- 4:40 - 5:05 Erika F. de Lima
Assigning Grammatical Relations with a Back-off Model
- 5:05 - 5:30 I. Dan Melamed
Automatic Discovery of Non-Compositional Compounds in Parallel Data

CONFERENCE PROGRAM

Saturday, August 2

- 8:00 - 8:30 **Registration and Continental Breakfast**
- 8:30 - 8:55 **Scott W. Bennett, Chinatsu Aone, and Craig Lovell**
Learning to Tag Multilingual Texts Through Observation
- 8:55 - 9:20 **Ellen Riloff and Jessica Shepherd**
A Corpus-Based Approach for Building Semantic Lexicons
- 9:20 - 9:45 **Roberto Basili, Gianluca De Rossi, and Maria Teresa Pazienza**
Inducing Terminology for Lexical Acquisition
- 9:45 - 10:15 **Break**
- 10:15 - 10:40 **Paul Thompson and Christopher C. Dozier**
Name Searching and Information Retrieval
- 10:40 - 11:05 **K.L. Kwok**
Lexicon Effects on Chinese Information Retrieval
- 11:05 - 11:25 **Break**
- 11:25 - 11:50 **Paola Merlo, Matthew W. Crocker, and Cathy Berthouzoz**
Attaching Multiple Prepositional Phrases: Generalized Backed-off Estimation
- 11:50 - 12:15 **Eric V. Siegel**
Learning Methods for Combining Linguistic Indicators to Classify Verbs
- 12:15 - 12:30 **SIGDAT Business Meeting**
- 12:30 - 1:45 **CATERED LUNCH**
- 1:45 - 2:45 **INVITED TALK, Michael Mauldin**
The Lycos System: Practical Information Retrieval
- 2:45 - 3:00 **Break**
- 3:00 - 3:25 **Andrew Kehler**
Probabilistic Coreference in Information Extraction
- 3:25 - 3:50 **Janyce Wiebe, Tom O'Hara, Kenneth McKeever, and Thorsten Öhrström-Sandgren**
An Empirical Approach to Temporal Reference Resolution
- 3:50 - 4:15 **Break**
- 4:15 - 4:40 **Ji Donghong and Huang Changning**
Word Sense Disambiguation Based on Structured Semantic Space
- 4:40 - 5:05 **Ted Pedersen and Rebecca Bruce**
Distinguishing Word Senses in Untagged Text
- 5:05 - 5:30 **Hwee Tou Ng**
Exemplar-Based Word Sense Disambiguation: Some Recent Improvements

TABLE OF CONTENTS

<i>A Linear Observed Time Statistical Parser Based on Maximum Entropy Models</i>	
Adwait Ratnaparkhi	1
<i>Global Thresholding and Multiple-Pass Parsing</i>	
Joshua Goodman	11
<i>An Efficient Distribution of Labor in a Two Stage Robust Interpretation Process</i>	
Carolyn Penstein Rosé and Alon Lavie	26
<i>Text Segmentation Using Exponential Models</i>	
Doug Beeferman, Adam Berger, and John Lafferty	35
<i>Detecting Subject Boundaries Within Text: A Language Independent Statistical Approach</i>	
Korin Richmond, Andrew Smith, and Einat Amitay	47
<i>Mistake-Driven Learning in Text Categorization</i>	
Ido Dagan, Yael Karov, and Dan Roth	55
<i>Tagging Grammatical Functions</i>	
Thorsten Brants, Wojciech Skut, and Brigitte Krenn	64
<i>On Aligning Trees</i>	
Jo Calder	75
<i>Aggregate and Mixed-order Markov Models for Statistical Language Processing</i>	
Lawrence Saul and Fernando Pereira	81
<i>Assigning Grammatical Relations with a Back-off Model</i>	
Erika F. de Lima	90
<i>Automatic Discovery of Non-Compositional Compounds in Parallel Data</i>	
I. Dan Melamed	97
<i>Learning to Tag Multilingual Texts Through Observation</i>	
Scott W. Bennett, Chinatsu Aone, and Craig Lovell	109
<i>A Corpus-Based Approach for Building Semantic Lexicons</i>	
Ellen Riloff and Jessica Shepherd	117
<i>Inducing Terminology for Lexical Acquisition</i>	
Roberto Basili, Gianluca De Rossi, and Maria Teresa Pazienza	125
<i>Name Searching and Information Retrieval</i>	
Paul Thompson and Christopher C. Dozier	134
<i>Lexicon Effects on Chinese Information Retrieval</i>	
K.L. Kwok	141
<i>Attaching Multiple Prepositional Phrases: Generalized Backed-off Estimation</i>	
Paola Merlo, Matthew W. Crocker, and Cathy Berthouzoz	149
<i>Learning Methods for Combining Linguistic Indicators to Classify Verbs</i>	
Eric V. Siegel	156
<i>Probabilistic Coreference in Information Extraction</i>	
Andrew Kehler	163
<i>An Empirical Approach to Temporal Reference Resolution</i>	
Janyce Wiebe, Tom O’Hara, Kenneth McKeever, and Thorsten Öhrström-Sandgren	174
<i>Word Sense Disambiguation Based on Structured Semantic Space</i>	
Ji Donghong and Huang Changning	187
<i>Distinguishing Word Senses in Untagged Text</i>	
Ted Pedersen and Rebecca Bruce	197
<i>Exemplar-Based Word Sense Disambiguation: Some Recent Improvements</i>	
Hwee Tou Ng	208

AUTHOR INDEX

Einat Amitay	47
Chinatsu Aone	109
Roberto Basili	125
Doug Beferman	35
Scott W. Bennett	109
Adam Berger	35
Cathy Berthouzoz	149
Thorsten Brants	64
Rebecca Bruce	197
Jo Calder	75
Huang Changning	187
Matthew W. Crocker	149
Ido Dagan	55
Erika F. de Lima	90
Gianluca De Rossi	125
Ji Donghong	187
Christopher C. Dozier	134
Joshua Goodman	11
Yael Karov	55
Andrew Kehler	163
Brigitte Krenn	64
K.L. Kwok	141
John Lafferty	35
Alon Lavie	26
Craig Lovell	109
Kenneth McKeever	174
I. Dan Melamed	97
Paola Merlo	149
Hwee Tou Ng	208
Tom O'Hara	174
Thorsten Öhrström-Sandgren	174
Maria Teresa Paziienza	125
Ted Pedersen	197
Fernando Pereira	81
Adwait Ratnaparkhi	1
Korin Richmond	47
Ellen Riloff	117
Carolyn Penstein Rosé	26
Dan Roth	55
Lawrence Saul	81
Jessica Shepherd	117
Eric V. Siegel	156
Wojciech Skut	64
Andrew Smith	47
Paul Thompson	134
Janyce Wiebe	174

