SPONSORS:

The Association for Computational Linguistics (ACL)
SIGDAT (ACL's SIG for Linguistic Data and Corpus-based Approaches to NLP)
SIGNLL (ACL's SIG for Natural Language Learning)
LEXIS-NEXIS, a Division of Reed Elsevier, Inc.

INVITED SPEAKERS:

Mark Liberman
Henry Kučera and Nelson Francis

ORGANIZERS:

David Yarowsky, Chair
Kenneth Church, Co-chair

PROGRAM COMMITTEE:

Susan Armstrong (ISCCO, Switzerland)
Walter Daelemans (ITK/KUB, Tilburg, Netherlands)
Marti Hearst (Xerox PARC, USA)
Chang-Ning Huang (Tsinghua University, China)
Pierre Isabelle (CITI, Canada)
Yuji Matsumoto (NAIST, Japan)
David Powers (Flinders University, Australia)
Philip Resnik (Sun Microsystems Laboratories, USA)
Dekai Wu (HKUST, Hong Kong)
Joe Zhou (LEXIS-NEXIS, USA)

ADDITIONAL REVIEWERS:

Yves Schabes (Mitsubishi Electric Research Labs, USA)
Julian Kupiec (Xerox PARC, USA)
Takehito Utsuro (NAIST, Japan)

FURTHER INFORMATION:

David Yarowsky
Dept. of Computer and Info. Science
University of Pennsylvania
200 S. 33rd St.
Philadelphia, PA 19104-6389 USA
email: yarowsky@unagi.cis.upenn.edu

Kenneth Church
Room 2B-421
AT&T Bell Laboratories
600 Mountain Ave.
Murray Hill, NJ 07974 USA
e-mail: kwc@research.att.com
WORKSHOP PROGRAM

8:45 - 8:50 Welcome
8:50 - 9:35 INVITED TALK (Mark Liberman)
9:35 - 9:50 Break
9:50 - 10:15 Eric Brill
Unsupervised Learning of Disambiguation Rules for Part of Speech Tagging
10:15 - 10:40 Carl de Marcken
Lexical Heads, Phrase Structure and the Induction of Grammar
10:40 - 11:05 Michael Collins and James Brooks
Prepositional Phrase Attachment through a Backed-off Model
11:05 - 11:15 Break
11:15 - 11:40 Andrew Golding
A Bayesian Hybrid Method for Context-sensitive Spelling Correction
11:40 - 12:05 Philip Resnik
Disambiguating Noun Groupings with Respect to Wordnet Senses
12:05 - 1:05 LUNCH
1:05 - 1:30 Dekai Wu
Trainable Coarse Bilingual Grammars for Parallel Text Bracketing
1:30 - 1:55 Lance Ramshaw and Mitch Marcus
Text Chunking using Transformation-Based Learning
1:55 - 2:05 Break
2:05 - 3:00 INVITED TALK (Henry Kučera and Nelson Francis)
3:00 - 3:10 Break
3:10 - 3:35 Fernando Pereira, Yoram Singer and Naftali Tishby
Beyond Word N-Grams
3:35 - 4:00 Jing-Shin Chang, Yi-Chung Lin and Keh-Yih Su
Automatic Construction of a Chinese Electronic Dictionary
4:00 - 4:10 Break
4:10 - 4:35 Kenneth Church and William Gale
Inverse Document Frequency (IDF): A Measure of Deviations from Poisson
4:35 - 5:00 Joe Zhou and Pete Dapkus
Automatic Suggestion of Significant Terms for a Predefined Topic
5:00 - 5:25 Ellen Riloff and Jay Shoem
Automatically Acquiring Conceptual Patterns without an Annotated Corpus
# TABLE OF CONTENTS

- **Unsupervised Learning of Disambiguation Rules for Part of Speech Tagging**
  - Eric Brill ................................................. 1

- **Lexical Heads, Phrase Structure and the Induction of Grammar**
  - Carl de Marcken ........................................... 14

- **Prepositional Phrase Attachment through a Backed-off Model**
  - Michael Collins and James Brooks ....................... 27

- **A Bayesian Hybrid Method for Context-sensitive Spelling Correction**
  - Andrew Golding ............................................. 39

- **Disambiguating Noun Groupings with Respect to Wordnet Senses**
  - Philip Resnik ............................................. 54

- **Trainable Coarse Bilingual Grammars for Parallel Text Bracketing**
  - Dekai Wu ................................................. 69

- **Text Chunking using Transformation-Based Learning**
  - Lance Ramshaw and Mitch Marcus ....................... 82

- **Beyond Word N-Grams**
  - Fernando Pereira, Yoram Singer and Naftali Tishby .......... 95

- **Automatic Construction of a Chinese Electronic Dictionary**
  - Jing-Shin Chang, Yi-Chung Lin and Keh-Yih Su ............ 107

- **Inverse Document Frequency (IDF): A Measure of Deviations from Poisson**
  - Kenneth Church and William Gale .......................... 121

- **Automatic Suggestion of Significant Terms for a Predefined Topic**
  - Joe Zhou and Pete Dapkus .................................. 131

- **Automatically Acquiring Conceptual Patterns without an Annotated Corpus**
  - Ellen Riloff and Jay Shoen .............................. 148

- **Development of a Partially Bracketed Corpus with Part-of-Speech Information Only**
  - Hsin-Hsi Chen and Yue-Shi Lee .......................... 162

- **Compiling Bilingual Lexicon Entries From a Non-Parallel English-Chinese Corpus**
  - Pascale Fung ............................................. 173

- **Automatic Evaluation and Uniform Filter Cascades for Inducing N-Best Translation Lexicons**
  - I. Dan Melamed ............................................ 184
AUTHOR INDEX

Eric Brill ................................................................. 1
James Brooks .......................................................... 27
Jing-Shin Chang ...................................................... 107
Hsin-Hsi Chen ......................................................... 162
Kenneth Church ...................................................... 121
Michael Collins ....................................................... 27
Pete Dapkus ............................................................ 131
Carl de Marcken ....................................................... 14
Pascale Fung ........................................................... 173
William Gale ........................................................... 121
Andrew Golding ...................................................... 39
Yue-Shi Lee ............................................................. 162
Yi-Chung Lin ........................................................... 107
Mitch Marcus .......................................................... 82
I. Dan Melamed ....................................................... 184
Fernando Pereira ..................................................... 95
Lance Ramshaw ...................................................... 82
Philip Resnik .......................................................... 54
Ellen Riloff ............................................................... 148
Jay Shoen ............................................................... 148
Yoram Singer ........................................................ 95
Keh-Yih Su ............................................................. 107
Naftali Tishby ......................................................... 95
Dekai Wu ............................................................... 69
Joe Zhou ............................................................... 131