

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

**Sponsored by:**

**The Association for Computational Linguistics**  
**Department of Computer and Information Science, University**  
**of Pennsylvania**

**Institute for Research in Cognitive Science, University of**  
**Pennsylvania**

**Lexis-Nexis, a Division of Reed Elsevier, Inc.**  
**SIGDAT, a Special Interest Group of the ACL**

**Edited by**

**Eric Brill**  
**and Kenneth Church**

**17-18 May 1996**  
**University of Pennsylvania**  
**Philadelphia, Pa. USA**



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

**Sponsored by:**

**The Association for Computational Linguistics**  
**Department of Computer and Information Science, University**  
**of Pennsylvania**  
**Institute for Research in Cognitive Science, University of**  
**Pennsylvania**  
**Lexis-Nexis, a Division of Reed Elsevier, Inc.**  
**SIGDAT, a Special Interest Group of the ACL**

**Edited by**

**Eric Brill**  
**and Kenneth Church**

**17-18 May 1996**  
**University of Pennsylvania**  
**Philadelphia, Pa. USA**



**Sponsors:**

Association for Computational Linguistics  
Department of Computer and Information Science, University of  
Pennsylvania  
Institute for Research in Cognitive Science, University of Pennsylvania  
Lexis-Nexis, a Division of Reed Elsevier, Inc.  
SIGDAT, a Special Interest Group of the ACL

**Organizers:**

Eric Brill (Johns Hopkins University)  
Kenneth Church (AT&T Research)

**Program Committee:**

Claire Cardie (Cornell University)  
Eugene Charniak (Brown University)  
Yuqing Gao (IBM)  
Marti Hearst (Xerox PARC)  
Gunnel Kallgren (Stockholm University)  
Raymond J. Mooney (University of Texas, Austin)  
Yoshinori Sagisaka (ATR)  
Geoffrey Sampson (Sussex University)  
David Yarowsky (Johns Hopkins University)  
Joe Zhou (Lexis-Nexis)

**For further information, contact:**

Eric Brill  
brill@cs.jhu.edu  
Department of Computer Science  
Johns Hopkins University  
Baltimore, Md. 21218-2694



# TABLE OF CONTENTS

<i>A Geometric Approach to Mapping Bitext Correspondence</i> I. Dan Melamed .....	1
<i>Parsing Chinese With an Almost-Context-Free Grammar</i> Xuanyin Xia and Dekai Wu .....	13
<i>Unsupervised Learning of Syntactic Knowledge: Methods and Measures</i> R. Basili, A. Marziali, M.T. Pazienza, P. Velardi .....	23
<i>Modeling Conversational Speech for Speech Recognition</i> Marie Meteer and Rukmini Iyer .....	33
<i>Automatic Extraction of New Words from Japanese Texts using Generalized Forward-Backward Search</i> Masaaki Nagata .....	48
<i>Better Language Models with Model Merging</i> Thorsten Brants .....	60
<i>Combining Hand-crafted Rules and Unsupervised Learning in Constraint-based Morphological Disambiguation</i> Kemal Oflazer and Gokhan Tur .....	69
<i>Comparative Experiments on Disambiguating Word Senses: An Illustration of the Role of Bias in Machine Learning</i> Raymond J. Mooney .....	82
<i>Apportioning Development Effort in a Probabilistic LR Parsing System Through Evaluation</i> John Carroll and Ted Briscoe .....	92
<i>The Measure of a Model</i> Rebecca Bruce, Janyce Wiebe, Ted Pedersen .....	101
<i>Automating Feature Set Selection for Case-Based Learning of Linguistic Knowledge</i> Claire Cardie .....	113
<i>Figures of Merit for Best-First Probabilistic Chart Parsing</i> Sharon A. Caraballo and Eugene Charniak .....	127
<i>A Maximum Entropy Model for Part-Of-Speech Tagging</i> Adwait Ratnaparkhi .....	133
<i>Efficient Algorithms for Parsing the DOP Model</i> Joshua Goodman .....	143

## AUTHOR INDEX

R. Basili.....	23
Thorsten Brants .....	60
Ted Briscoe .....	92
Rebecca Bruce .....	101
Sharon A. Caraballo .....	127
Claire Cardie .....	113
John Carroll .....	92
Eugene Charniak .....	127
Joshua Goodman .....	143
Rukmini Iyer .....	33
A. Marziali .....	23
I. Dan Melamed .....	1
Marie Meter .....	33
Raymond J. Mooney .....	82
Masaaki Nagata .....	48
Kemal Oflazer .....	69
M.T. Paziienza .....	23
Ted Pedersen .....	101
Adwait Ratnaparkhi .....	133
Gokhan Tur .....	69
P. Velardi .....	23
Janyce Wiebe .....	101
Dekai Wu .....	13
Xuanyin Xia .....	13