

# Speech and Natural Language

**Proceedings of a workshop  
held at  
Harriman, New York  
February 23-26, 1992**

**Sponsored by:  
Defense Advanced Research Projects Agency  
Software and Intelligent Systems Technology Office**

This document contains copies of reports prepared for the DARPA Speech and Natural Language Workshop. Included are reports from DARPA/SISTO sponsored programs and other materials prepared for use at the workshop.

**APPROVED FOR PUBLIC RELEASE  
DISTRIBUTION UNLIMITED**

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the United States Government.

Distributed by:  
Morgan Kaufmann Publishers, Inc.  
2929 Campus Drive, Suite 260  
San Mateo, California 94403  
ISBN 1-55860-272-0  
Printed in the United States of America

# TABLE OF CONTENTS

	Page
<b>Author Index</b> .....	x
<b>Overview of the Workshop, Mitchell Marcus, General Chair</b> .....	3
<b>SESSION 1: Evaluating Spoken Language Systems</b>	
<b>Session Summary, James Allen, Chair</b> .....	5
<b>Multi-Site Data Collection For a Spoken Language Corpus</b>	
MADCOW (Contact: L. Hirschman); MIT Laboratory for Computer Science.....	7
<b>DARPA February 1992 ATIS Benchmark Test Results</b>	
D. Pallett, N. Dahlgren, J. Fiscus, W. Fisher, J. Garofolo and B. Tjaden; National Institute for Standards and Technology.....	15
<b>Experiments in Evaluating Interactive Spoken Language Systems</b>	
J. Polifroni, L. Hirschman, S. Seneff and V. Zue; MIT Laboratory for Computer Science.....	28
<b>Subject-Based Evaluation Measures for Interactive Spoken Language Systems</b>	
P. Price, L. Hirschman, E. Shriberg and E. Wade; SRI International, MIT Laboratory for Computer Science, UC Berkeley and Stanford University.....	34
<b>SESSION 2: Spoken Language Systems II</b>	
<b>Session Summary, Wayne Ward, Chair</b> .....	41
<b>Spontaneous Speech Collection for the ATIS Domain with an Aural User Feedback Paradigm</b>	
C. Pao and J. Wilpon; AT&T Bell Laboratories.....	43
<b>Human-Machine Problem Solving Using Spoken Language Systems (SLS): Factors Affecting Performance and User Satisfaction</b>	
E. Shriberg, E. Wade and P. Price; SRI International.....	49
<b>Spoken Language Processing in the Framework of Human-Machine Communication at LIMSI</b>	
J. Mariani; LIMSI-CNRS.....	55
<b>Experiences Collecting Genuine Spoken Enquiries Using WOZ Techniques</b>	
R. Moore and A. Morris; RSRE Speech Research Unit.....	61
<b>SESSION 3: Spoken Language Systems III</b>	
<b>Session Summary, John Makhoul, Chair</b> .....	65

	<b>Page</b>
<b>Progress Report on the Chronus System: ATIS Benchmark Results</b> R. Pieraccini, E. Tzoukermann, Z. Gorelov, E. Levin and C.-H. Lee; AT&T Bell Laboratories.....	67
<b>BBN BYBLOS and HARC February 1992 ATIS Benchmark Results</b> F. Kubala, C. Barry, M. Bates, R. Bobrow, P. Fung, R. Ingria, J. Makhoul, L. Nguyen, R. Schwartz, and D. Stallard; BBN Systems and Technologies.....	72
<b>Speech Recognition in Open Tasks</b> W. Ward, S. Issar, X. Huang, H. Hon, M. Hwang, S. Young, M. Matessa, F. Liu and R. Stern; Carnegie Mellon University.....	78
<b>The MIT ATIS System: February 1991 Progress Report</b> V. Zue, J. Glass, D. Goddeau, D. Goodine, L. Hirschman, M. Phillips, J. Polifroni, and S. Seneff; MIT Laboratory for Computer Science.....	84
<b>Recent Improvements and Benchmark Results for the Paramax ATIS System</b> L. Norton, D. Dahl and M. Linebarger; Paramax.....	89
<b>SRI International February 1992 ATIS Benchmark Test Results</b> D. Appelt and E. Jackson; SRI International.....	95
 <b>SESSION 4: Statistical Language Modeling</b>	
<b>Session Summary, Aravind Joshi, Chair</b> .....	101
 <b>Adaptive Language Modeling Using Minimum Discriminant Estimation</b> S. Della Pietra, V. Della Pietra, R.L. Mercer and S. Roukos; IBM T.J. Watson Research Center.....	103
<b>Improvements in Stochastic Language Modeling</b> R. Rosenfeld and X. Huang; Carnegie Mellon University.....	107
<b>A Simple Rule-Based Part of Speech Tagger</b> E. Brill; University of Pennsylvania.....	112
<b>Decision Tree Models Applied to the Labeling of Text with Parts-of-Speech</b> E. Black, F. Jelinek, J. Lafferty, R. Mercer and S. Roukos; IBM T. J. Watson Research Center.....	117
<b>Inside-Outside Reestimation from Partially Bracketed Corpora</b> F. Pereira and Y. Schabes; AT&T Bell Laboratories and University of Pennsylvania.....	122
<b>Probabilistic Prediction and Picky Chart Parsing</b> D. Magerman and C. Weir; Stanford University and Paramax Systems Corporation.....	128
<b>Towards History-based Grammars: Using Richer Models for Probabilistic Parsing</b> E. Black, F. Jelinek, J. Lafferty, D. Magerman, R. Mercer and S. Roukos; IBM T.J. Watson Research Center.....	134
<b>Stochastic Tree-Adjoining Grammars</b> Y. Schabes; University of Pennsylvania.....	140
<b>Parameter Estimation for Constrained Context-Free Language Models</b> K. Mark, M. Miller, U. Grenander and S. Abney; Washington University, Brown University, and Bell Communications Research.....	146

	<b>Page</b>
<b>An Analogical Parser for Restricted Domains</b> D. Hindle; AT&T Bell Laboratories.....	150
<b>Automatically Acquiring Phrase Structure Using Distributional Analysis</b> E. Brill and M. Marcus; University of Pennsylvania.....	155
 <b>SESSION 5a: Acoustic Modeling</b>	
<b>Session Summary, Hy Murveit, Chair</b> .....	161
<b>Recent Topics in Speech Recognition Research at NTT Laboratories</b> S. Furui, K. Shikano, S. Matsunaga, T. Matsuoka, S. Takahashi and T. Yamada; NTT Human Interface Laboratories.....	162
<b>Vocabulary and Environment Adaptation in Vocabulary-Independent Speech Recognition</b> H.-W. Hon and K.-F. Lee; Carnegie Mellon University and Apple Computer, Inc.....	168
<b>Subphonetic Modeling for Speech Recognition</b> M.-Y. Hwang and X. Huang; Carnegie Mellon University.....	174
<b>Improving State-of-the-Art Continuous Speech Recognition Systems Using the N-Best Paradigm with Neural Networks</b> S. Austin, G. Zavaliagkos, J. Makhoul and R. Schwartz; BBN Systems and Technologies.....	180
<b>MAP Estimation of Continuous Density HMM: Theory and Applications</b> J.-L. Gauvain and C.-H. Lee; AT&T Bell Laboratories.....	185
<b>Minimizing Speaker Variation Effects for Speaker-Independent Speech Recognition</b> X. Huang; Carnegie Mellon University.....	191
<b>Recognition Using Classification and Segmentation Scoring</b> O. Kimball, M. Ostendorf and R. Rohlicek; Boston University.....	197
 <b>SESSION 5b: Information Retrieval</b>	
<b>Session Summary, Donna Harman, Chair</b> .....	203
<b>Information Retrieval Using Robust Natural Language Processing</b> T. Strzalkowski; New York University.....	206
<b>Feature Selection and Feature Extraction for Text Categorization</b> D. Lewis; University of Chicago.....	212
<b>Inferencing in Information Retrieval</b> A. McCray; National Library of Medicine.....	218
<b>Classifying Texts Using Relevancy Signatures</b> E. Riloff and W. Lehnert; University of Massachusetts-Amherst.....	224

**SESSION 6: Lexicon and Lexical Semantics**

**Session Summary, Paul Jacobs, Chair**.....231

**One Sense per Discourse**

W. Gale, K. Church and D. Yarowsky; AT&T Bell Laboratories.....233

**Lexical Disambiguation Using Simulated Annealing**

J. Cowie, L. Guthrie and J. Guthrie; New Mexico State University.....238

**The Acquisition of Lexical Semantic Knowledge from Large Corpora**

J. Pustejovsky; Brandeis University.....243

**SESSION 7: Demonstrations and Videos**

**Session Summary, Victor Zue, Chair**.....249

**BBN Real-Time Speech Recognition Demonstrations**

S. Austin, R. Bobrow, D. Ellard, R. Ingria, J. Makhoul, L. Nguyen, P. Peterson,  
P. Placeway and R. Schwartz; BBN Systems and Technologies.....250

**SESSION 8a: Machine Translation**

**Session Summary, Jerry Hobbs, Chair**.....253

**Interaction Between Structural Changes in Machine Translation**

S. Kinoshita, J. Phillips and J. Tsujii; University of Manchester Institute of Science and Technology.....255

**Approximating an Interlingua in a Principled Way**

E. Hovy and S. Nirenburg; Information Sciences Institute and Carnegie Mellon University.....261

**Dividing and Conquering Long Sentences in a Translation System**

P. Brown, S. Della Pietra, V. Della Pietra, R. Mercer and S. Mohanty;  
IBM T.J. Watson Research Center.....267

**SESSION 8b: Robust Speech Processing**

**Session Summary, Jordan Cohen, Chair**.....273

**Multiple Approaches to Robust Speech Recognition**

R. Stern, F.-H. Liu, Y. Ohshima, T. Sullivan and A. Acero; Carnegie Mellon University.....274

**Reduced Channel Dependence for Speech Recognition**

H. Murveit, J. Butzberger and M. Weintraub; SRI International.....280

	<b>Page</b>
<b>Experimental Results for Baseline Speech Recognition Performance using Input Acquired from a Linear Microphone Array</b> H. Silverman, S. Kirtman, J. Adcock and P. Meuse; Brown University.....	285
<b>Phonetic Classification on Wide-Band and Telephone Quality Speech</b> B. Chigier; NYNEX Science & Technology.....	291
 <b>SESSION 9: Natural Language Processing</b>	
<b>Session Summary, Kathleen McKeown, Chair</b> .....	297
<b>A Relaxation Method for Understanding Spontaneous Utterances</b> S. Seneff; MIT Laboratory for Computer Science.....	299
<b>Fragment Processing in the DELPHI System</b> D. Stallard and R. Bobrow; BBN Systems and Technologies.....	305
<b>Syntactic/Semantic Coupling in the BBN DELPHI System</b> R. Bobrow, R. Ingria and D. Stallard; BBN Systems and Technologies.....	311
<b>A New Approach to Text Understanding</b> R. Weischedel, D. Ayuso, S. Boisen, H. Fox and R. Ingria; BBN Systems and Technologies.....	316
<b>Neal-Montgomery NLP System Evaluation Methodology</b> S. Walter; Rome Laboratory.....	323
 <b>SESSION 10a: Acoustic Modeling</b>	
<b>Session Summary, George Doddington, Chair</b> .....	327
<b>Rapid Match Training for Large Vocabularies</b> L. Gillick, B. Peskin and R. Roth; Dragon Systems.....	328
<b>An A* Algorithm for Very Large Vocabulary Continuous Speech Recognition</b> P. Kenny, R. Hollan, G. Boulianne, H. Garudadri, M. Lennig and D. O'Shaughnessy; INRS-Telecommunications.....	333
<b>Spontaneous Speech Effects in Large Vocabulary Speech Recognition Applications</b> J. Butzberger, H. Murveit, E. Shriberg and P. Price; SRI International.....	339
<b>Speaker-Independent Phone Recognition Using BREF</b> J.-L. Gauvain and L. Lamel; LIMSI-CNRS.....	344
 <b>SESSION 10b: Core NL Lexicon and Grammar</b>	
<b>Session Summary, Mark Liberman, Chair</b> .....	351

	<b>Page</b>
<b>A National Resource Grammar</b> J. Hobbs; SRI International.....	352
 <b>SESSION 11: Continuous Speech Recognition and Evaluation I</b>	
<b>Session Summary, Cliff Weinstein, Chair</b> .....	355
 <b>The Design for the Wall Street Journal-based CSR Corpus</b> D. Paul and J. Baker; MIT Lincoln Laboratory and Dragon Systems.....	357
<b>CSR Corpus Development</b> G. Doddington; SRI International.....	363
<b>Collection and Analyses of WSJ-CSR Data at MIT</b> M. Phillips, J. Glass, J. Polifroni, and V. Zue; MIT Laboratory for Computer Science.....	367
<b>Spontaneous Speech Collection for the CSR Corpus</b> J. Bernstein and D. Danielson; SRI International.....	373
 <b>SESSION 12: Continuous Speech Recognition and Evaluation II</b>	
<b>Session Summary, Cliff Weinstein, Chair</b> .....	379
 <b>DARPA February 1992 Pilot Corpus CSR "Dry Run" Benchmark Test Results</b> D. Pallett; National Institute for Standards and Technology.....	382
<b>Large Vocabulary Recognition of Wall Street Journal Sentences at Dragon Systems</b> J. Baker, J. Baker, P. Bamberg, K. Bishop, L. Gillick, V. Helman, Z. Huang, Y. Ito, S. Lowe, B. Peskin, R. Roth and F. Scattone; Dragon Systems.....	387
<b>Applying Sphinx-II to the DARPA Wall Street Journal CSR Task</b> F. Alleva, H. Hon, X. Huang, M. Hwang, R. Rosenfeld and R. Weide; Carnegie Mellon University.....	393
<b>The Lincoln Large-Vocabulary HMM CSR</b> D. Paul; MIT Lincoln Laboratory.....	399
<b>An Efficient A* Stack Decoder Algorithm for Continuous Speech Recognition with a Stochastic Language Model</b> D. Paul; MIT Lincoln Laboratory.....	405
<b>Performance of SRI's Decipher Speech Recognition System on DARPA's CSR Task</b> H. Murveit, J. Butzberger and M. Weintraub; SRI International.....	410
 <b>SESSION 13: Prosody</b>	
<b>Session Summary, Patti Price and Julia Hirschberg, Chairs</b> .....	415
 <b>Automatic Detection and Correction of Repairs in Human-Computer Dialog</b> E. Shriberg, J. Bear and J. Dowding; SRI International.....	419



	<b>Page</b>
<b>Prosodic Structure, Performance Structure and Phrase Structure</b> S. Abney; Bell Communications Research.....	425
<b>Probabilistic Parse Scoring Based on Prosodic Phrasing</b> N. Veilleux and M. Ostendorf; Boston University.....	429
<b>Towards Using Prosody in Speech Recognition/Understanding Systems: Differences Between Read and Spontaneous Speech</b> K. Silverman, E. Blaauw, J. Spitz and J. Pitrelli; NYNEX Science & Technology, Inc.....	435
<b>Intonational Features of Local and Global Discourse Structure</b> J. Hirschberg and B. Grosz; AT&T Bell Laboratories and Harvard University.....	441
 <b>SESSION 14: Government Panel</b>	
<b>Session Summary, Charles Wayne, Chair</b> .....	447
<b>Language and Speech Meets the HPCC Grand Challenge</b> Y.-T. Chien; National Science Foundation.....	449
 <b>POSTERS</b>	
<b>Comparison of Auditory Models for Robust Speech Recognition</b> C. Jankowski and R. Lippmann; MIT Lincoln Laboratory.....	453
<b>Weight Estimation for N-Best Rescoring</b> A. Kannan, M. Ostendorf, and J.R. Rohlicek; Boston University.....	455
<b>Augmenting Relevancy Signatures with Slot Filler Data</b> E. Riloff and W. Lehnert; University of Massachusetts-Amherst.....	457
<b>Language Understanding Research at Paramax</b> D. Dahl, C. Weir, S. Taylor, L. Norton, M. Linebarger and M. Lipschutz; Paramax.....	459
<b>SITE REPORTS</b> .....	461

## AUTHOR INDEX

Abney, S. ....	146, 425	Della Pietra, V. ....	103, 267
Acero, A. ....	274	Doddington, G. ....	327, 363
Adcock, J. ....	285	Dowding, J. ....	419
Allen, J. ....	5	Ellard, D. ....	250
Alleva, F. ....	393	Fiscus, J. ....	15
Appelt, D. ....	95	Fisher, W. ....	15
Austin, S. ....	180, 250	Fox, H. ....	316
Ayuso, D. ....	316	Fung, P. ....	72
Baker, J. K. ....	387	Furui, S. ....	162
Baker, J. M. ....	357, 387, 471	Gale, W. ....	233
Bamberg, P. ....	387	Garofolo, J. ....	15
Barry, C. ....	72	Garudadri, H. ....	333
Bates, M. ....	72, 463	Gauvain, J.-L. ....	67, 185, 344
Bear, J. ....	419	Gillick, L. ....	328, 387, 471
Bernstein, J. ....	373, 484, 485	Glass, J. ....	84, 367
Bishop, K. ....	387	Goddeau, D. ....	84
Blaauw, E. ....	435	Goodine, D. ....	84
Black, E. ....	117, 134	Gorelov, Z. ....	67
Bobrow, R. ....	72, 250, 305, 311	Grenander, U. ....	146
Boisen, S. ....	316	Grishman, R. ....	482
Boulianne, G. ....	333	Grosz, B. ....	441
Brill, E. ....	112, 155	Guthrie, J. ....	238
Brown, P. ....	267	Guthrie, L. ....	238
Butzberger, J. ....	280, 339, 410	Harman, D. ....	203
Chien, Y.-T. ....	449	Helman, V. ....	387
Chigier, B. ....	291	Hindle, D. ....	150
Church, K. ....	233	Hirschberg, J. ....	415, 441
Cohen, J. ....	273	Hirschman, L. ....	7, 28, 34, 84, 474
Cowie, J. ....	238, 479	Hobbs, J. ....	253, 352, 488
Dahl, D. ....	89, 459	Hollan, R. ....	333
Dahlgren, N. ....	15	Hon, H.-W. ....	78, 168, 393
Danielson, D. ....	373, 485	Hovy, E. ....	261, 491
Della Pietra, S. ....	103, 267	Huang, X. ....	78, 107, 174, 191, 393

Huang, Z. ....	387	Mariani, J. ....	55
Hunicke-Smith, K. ....	484	Mark, K. ....	146
Hwang, M.-Y. ....	78, 174, 393	Matessa, M. ....	78
Ingria, R. ....	72, 250, 311, 316	Matsunaga, S. ....	162
Issar, S. ....	78	Matsuoka, T. ....	162
Ito, Y. ....	387	McCray, A. ....	218
Jackson, E. ....	95	McKeown, K. ....	297, 470
Jacobs, P. ....	231	Mercer, R. ....	103, 117, 134, 267
Jankowski, C. ....	453	Meuse, P. ....	285
Jelinek, F. ....	117, 134	Miller, G. ....	483
Joshi, A. ....	101, 490	Miller, M. ....	146
Kannan, A. ....	455	Mohanty, S. ....	267
Kenny, P. ....	333	Moore, R. ....	61
Kimball, O. ....	197	Moore, R. C. ....	487
Kinoshita, S. ....	255	Morris, A. ....	61
Kirtman, S. ....	285	Murveit, H. ....	161, 280, 339, 410
Kubala, F. ....	72	Nguyen, L. ....	72, 250
Lafferty, J. ....	117, 134	Nirenburg, S. ....	261
Lamel, L. ....	344	Norton, L. ....	89, 459
Lee, C.-H. ....	67, 185	O'Shaughnessy, D. ....	333
Lee, K.-F. ....	168	Ohshima, Y. ....	274
Lehnert, W. ....	224, 457, 489	Ostendorf, M. ....	197, 429, 455, 466, 467
Lennig, M. ....	333	Pallett, D. ....	15, 382, 477
Levin, E. ....	67	Pao, C. ....	43
Lewis, D. ....	212	Passonneau, R. ....	470
Lieberman, M. ....	351	Paul, D. ....	357, 399, 405
Linebarger, M. ....	89, 459	Pereira, F. ....	122
Lippmann, R. ....	453, 475	Peskin, B. ....	328, 387
Lipschutz, M. ....	459	Peterson, P. ....	250
Litman, D. ....	470	Phillips, J. ....	255
Liu, F.-H. ....	78, 274	Phillips, M. ....	84, 367
Lowe, S. ....	387	Pieraccini, R. ....	67
Magerman, D. ....	128, 134	Pitrelli, J. ....	435
Makhoul, J. ....	65, 72, 180, 250, 463, 464	Placeway, P. ....	250
Marcus, M. ....	3, 155, 490	Polifroni, J. ....	28, 84, 367

Price, P. ....	34, 49, 339, 415, 466, 487	Webber, B. ....	490
Pustejovsky, J. ....	243, 479	Weide, R. ....	393
Reddy, R. ....	469	Weinstein, C. ....	355, 379, 476
Riloff, E. ....	224, 457	Weintraub, M. ....	280, 410, 486
Rohlicek, R. ....	197, 455, 467	Weir, C. ....	128, 459
Rosenfeld, R. ....	107, 393	Weischedel, R. ....	316, 465
Roth, R. ....	328, 387, 471	Wilks, Y. ....	479, 480, 481
Roukos, S. ....	103, 117, 134	Wilpon, J. ....	43
Scattone, F. ....	387	Yamada, T. ....	162
Schabes, Y. ....	122, 140	Yarowsky, D. ....	233
Schwartz, R. ....	72, 180, 250, 464	Young, S. ....	78
Seneff, S. ....	28, 84, 299	Zavaliagkos, G. ....	180
Shikano, K. ....	162	Zue, V. ....	28, 84, 249, 367, 474
Shore, J. ....	472		
Shriberg, E. ....	34, 49, 339, 419		
Silverman, H. ....	285, 468		
Silverman, K. ....	435		
Spitz, J. ....	435		
Stallard, D. ....	72, 305, 311		
Steedman, M. ....	490		
Stern, R. ....	78, 274		
Strzalkowski, T. ....	206		
Sullivan, T. ....	274		
Sundheim, B. ....	478		
Takahashi, S. ....	162		
Talkin, D. ....	473		
Taylor, S. ....	459		
Tjaden, B. ....	15		
Tsujii, J. ....	255		
Tzoukermann, E. ....	67		
Veilleux, N. ....	429		
Wade, E. ....	34, 49		
Walter, S. ....	323		
Ward, W. ....	41, 78		
Wayne, C. ....	447		