

HUMAN LANGUAGE TECHNOLOGY

**Proceedings of a workshop held at
Plainsboro, New Jersey
March 21-24, 1993**

**Sponsored by:
Advanced Research Projects Agency**

This document contains copies of reports prepared for the ARPA Human Language Technology Workshop. Included are reports from ARPA 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 Advanced research Projects Agency or the United States Government

Distributed by:
Morgan Kaufmann Publishers, Inc.
340 Pine Street, 6th Floor
San Francisco, CA 94104
ISBN 1-55860-324-7
Printed in the United States of America

TABLE OF CONTENTS

	Page	
Author Index	ix	
Overview of the ARPA Human Language Technology Workshop, Madeleine Bates, Chairperson	3	
 SESSION 1: Spoken Language Systems		
Session Summary, Alexander Rudnicky, Chair	5	
 Benchmark Tests for the DARPA Spoken Language Program David Pallett, Johathan Fiscus, William Fisher and John Garofolo; National Institute of Standards and Technology.....		7
 Multi-Site Data Collection and Evaluation in Spoken Language Understanding L. Hirschman, M. Bates, D. Dahl, W. Fisher, J. Garofolo, D. Pallett, K. Hunicke-Smith, P. Price, A. Rudnicky, and E. Tzoukermann.....		19
 The HCRC Map Task Corpus: Natural Dialogue for Speech Recognition Henry S. Thompson, Anne Anderson, Ellen Gurman Bard, Gwyneth Doherty-Sneddon, Alison Newlands, Cathy Sotillo; University of Edinburgh.....		25
 A Portable Approach to Last Resort Parsing and Interpretation Marcia C. Linebarger, Lewis M. Norton and Deborah A. Dahl; Paramax Systems Corporation.....		31
 The Semantic Linker - A New Fragment Combining Method David Stallard; Rusty Bobrow, BBN Systems and Technologies.....		37
 Gemini: A Natural Language System for Spoken-Language Understanding John Dowding, Jean Mark Gawron, Doug Appelt, John Bear, Lyn Cherny, Robert Moore and Doug Moran; SRI International.....		43
 A Bilingual VOYAGER System J. Glass, D. Goodine, M. Phillips, S. Sakai, S. Seneff and V. Zue; Massachusetts Institute of Technology.....		49
 SESSION 2: Invited Overviews		
Session Summary, Madeleine Bates, Chair	55	
 Survey of the Message Understanding Conferences Beth Sundheim; NRaD, Nancy Chinchor; Science Applications International Corporation.....		56
 Overview of TREC-1 Donna Harman; National Institute of Standards and Technology.....		61
 SESSION 3: Continuous Speech Recognition		
Session Summary, Douglas B. Paul, Chair	67	

	Page
Efficient Cepstral Normalization for Robust Speech Recognition Fu-Hua Liu, Richard M. Stern, Xuedong Huang, Alejandro Acero; Carnegie Mellon University.....	69
Comparative Experiments on Large Vocabulary Speech Recognition Richard Schwartz, Tasos Anastasakos, Francis Kubala, John Makhoul, Long Nguyen, George Zavalagkos ; BBN Systems and Technologies.....	75
An Overview of the SPHINX-II Speech Recognition System Xuedong Huang, Fileno Alleva, Mei-Yuh Hwang, Ronald Rosenfeld; Carnegie Mellon University.....	81
Progressive-Search Algorithms for Large-Vocabulary Speech Recognition Hy Murveit, John Butzberger, Vassilios Digalakis, Mitch Weintraub; SRI International.....	87
Search Algorithms for Software-Only Real-Time Recognition with Very Large Vocabularies Long Nguyen, Richard Schwartz, Francis Kubala, Paul Placeway; BBN Systems and Technologies.....	91
Identification of Non-Linguistic Speech Features Jean-Luc Gauvain and Lori F. Lamel; LIMSI-CNRS.....	96
On the Use of Tied-Mixture Distributions Owen Kimball and Mari Ostendorf; Boston University.....	102
Adaptive Language Modeling Using the Maximum Entropy Principle Raymond Lau, Ronald Rosenfeld, Salim Roukos; IBM Research Division.....	108
Improved Keyword-Spotting Using SRI's DECIPHER (TM) Large-Vocabulary Speech- Recognition System Mitchel Weintraub; SRI International.....	114
Topic and Speaker Identification via Large Vocabulary Continuous Speech Recognition Barbara Peskin, Larry Gillick, Yoshiko Ito, Stephen Lowe, Robert Roth, Francesco Scattone, James Baker, Janet Baker, John Bridle, Melvyn Hunt, Jeremy Orloff; Dragon Systems, Inc.	119
 SESSION 4: Natural Language	
Session Summary, Robert C. Moore, Chair.....	125
Heuristics for Broad-Coverage Natural Language Parsing Michael C. McCord; IBM T. J. Watson Research Center.....	127
FASTUS: A System for Extracting Information from Text Jerry R. Hobbs, Douglas Appelt, John Bear, David Israel, Megumi Kameyama, and Mabry Tyson; SRI International.....	133
Interpreting Temporal Adverbials Chung Hee Hwang & Lenhart K. Schubert; University of Rochester.....	138
The Murasaki Project: Multilingual Natural Language Understanding Chinatsu Aone, Hatté Blejer, Sharon Flank, Douglas McKee, Sandy Shinn; Systems Research and Applications.....	144
Validation of Terminological Inference in an Information Extraction Task Marc Vilain; The MITRE Corporation.....	150

SESSION 5: Discourse

Session Summary, Jerry R. Hobbs, Chair.....157

Development, Implementation and Testing of a Discourse Model for Newspaper Texts
Elizabeth D. Liddy; Syracuse University, Kenneth McVearry; Coherent Research, Inc., Woojin Paik; Syracuse University, Edmund Yu; Syracuse University, Mary McKenna; Syracuse University.....159

Indexing and Exploiting a Discourse History to Generate Context-Sensitive Explanations
Johanna D. Moore; University of Pittsburgh.....165

Generic Plan Recognition for Dialogue Systems
George Ferguson, James F. Allen; University of Rochester.....171

Efficient Collaborative Discourse: A Theory and Its Implementation
Alan W. Biermann, Curry I. Guinn, D. Richard Hipp, Ronnie W. Smith; Duke University.....177

SESSION 6: Machine Translation

Session Summary, Alex Waibel, Chair.....183

Building a Large Ontology for Machine Translation
Kevin Knight; USC/Information Sciences Institute.....185

LINGSTAT: An Interactive, Machine-Aided Translation System
Jonathan Yamron, James Baker, Paul Bamberg, Haakon Chevalier, Taiko Dietzel, John Elder, Frank Kampmann, Mark Mandel, Linda Manganaro, Todd Margolis, Elizabeth Steele; Dragon Systems, Inc.....191

An MAT Tool and Its Effectiveness
Robert Frederking, Dean Grannes, Peter Cousseau, Sergei Nirenburg; Carnegie Mellon University.....196

But Dictionaries are Data Too
Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Meredith J. Goldsmith, Jan Hajic, Robert L. Mercer, and Surya Mohanty; IBM T. J. Watson Research Center.....202

Evaluation of Machine Translation
John S. White; PRC, Theresa A. O'Connell; PRC, Lynn M. Carlson; DoD.....206

Recent Advances in Janus: A Speech Translation System
M. Woszczyna, N. Coccaro, A. Eisele, A. Lavie, A. McNair, T. Polzin, I. Rogina, C. P. Rose, T. Sloboda, M. Tomita, J. Tsutsumi, N. Waibel, A. Waibel, W. Ward; Carnegie Mellon University.....211

A Speech to Speech Translation System Built from Standard Components
Manny Rayner, Hiyan Alshawi; SRI International, Cambridge, UK, Ivan Bretan; SICS, David Carter; SRI International, Cambridge, UK, Vassilios Digalakis; SRI International, Menlo Park, CA, Bjorn Gamback; SICS, Jann Kaja; Telia Research, Jussi Karlgren; SICS, Bertil Lyberg; Telia Research, Steve Pulman; SRI International, Cambridge, UK, Patti Price; SRI International, Menlo Park, CA; and Christer Samuelsson; SICS.....217

SESSION 7: Demonstrations

Chair: Hy Murveit; SRI International.....223

	Page
SESSION 8: Statistical Language Modeling	
Session Summary, Mitchell Marcus, Chair.....	225
Example-Based Correction of Word Segmentation and Part of Speech Labelling	
Tomoyoshi Matsukawa, Scott Miller, Ralph Weischedel; BBN Systems and Technologies.....	227
Measures and Models for Phrase Recognition	
Steven Abney; Bell Communications Research.....	233
Automatic Grammar Induction and Parsing Free Text: A Transformation-Based Approach	
Eric Brill; University of Pennsylvania.....	237
Prediction of Lexicalized Tree Fragments in Text	
Donald Hindle; AT&T Bell Laboratories.....	243
Hypothesizing Word Association from Untagged Text	
Tomoyoshi Matsukawa; BBN Systems and Technologies.....	248
Smoothing of Automatically Generated Selectional Constraints	
Ralph Grishman and John Sterling; New York University.....	254
Corpus-Based Statistical Sense Resolution	
Claudia Leacock; Princeton University, Geoffrey Towell, Ellen Voorhees; Siemens University.....	260
One Sense Per Collocation	
David Yarowsky; University of Pennsylvania.....	266
Augmenting Lexicons Automatically: Clustering Semantically Related Adjectives	
Kathleen McKeown and Vasileios Hatzivassiloglou; Columbia University.....	272
Semantic Classes and Syntactic Ambiguity	
Philip Resnik; University of Pennsylvania.....	278
SESSION 9: Government Panel	
Session Summary, Carol J. Van Ess-Dykema, Chair.....	285
Projected Government Needs in Human Language Technology and the Role of Researchers in Meeting Them	
Helen M. Gigley; Naval Research Laboratory.....	287
Language Research Sponsored by ONR	
Susan Chipman; Office of Naval Research.....	290
Technology Transfer: Problems and Prospects	
Jesse Fussell; Department of Defense.....	295
SESSION 10: The Lexicon	
Session Summary, Ralph Grishman, Chair.....	299

	Page
The Complex Syntax Project	
Ralph Grishman, Catherine Macleod, Susanne Wolff; New York University.....	300
A Semantic Concordance	
George A. Miller, Claudia Leacock, Randee Teng, Ross Bunker; Princeton University.....	303
Intepretion of Proper Nouns for Information Retrieval	
Woojin Paik, Elizabeth D. Liddy, Edmund Yu, Mary McKenna; Syracuse University.....	309
SESSION 11: Prosody	
Session Summary, Mari Ostendorf, Chair.....	315
On Customizing Prosody in Speech Synthesis: Names and Addresses as a Case in Point	
Kim Silverman; NYNEX Science and Technology, Inc.....	317
Quantitative Modeling of Segmental Duration	
Jan P. H. Van Santen; AT&T Bell Laboratories.....	323
A Speech-First Model for Repair Detection and Correction	
Christine Nakatani; Harvard University, Julia Hirschberg; AT&T Bell Laboratories	329
Prosody/Parse Scoring and Its Application in ATIS	
N. M. Veilleux and M. Ostendorf; Boston University.....	335
Perceived Prosodic Boundaries and Their Phonetic Correlates	
Rene Collier, Jan Roelof de Pijper and Angelien Sanderman; Institute for Perception Research.....	341
SESSION 12: Document Retrieval and Text Retrieval	
Session Summary, Karen Sparck Jones, Chair.....	347
The Importance of Proper Weighting Methods	
Chris Buckley; Cornell University.....	349
Query Processing for Retrieval from Large Text Bases	
John Broglio and W. Bruce Croft; University of Massachusetts.....	353
An Overview of DR-LINK and Its Approach to Document Filtering	
Elizabeth D. Liddy, Woojin Paik, Edmund S. Yu; Syracuse University and Kenneth A. McVeary; Coherent Research, Inc.	358
SESSION 13: New Directions	
Session Summary, Ralph Weischedel, Chair.....	363
Mode Preference in a Simple Data-Retrieval Task	
Alexander I. Rudnicky; Carnegie Mellon University.....	364

	Page
A Simulation-Based Research Strategy for Designing Complex NL Systems Sharon Oviatt, Philip Cohen, Michelle Wang, Jeremy Gaston; SRI International.....	370
Speech and Text-Image Processing in Documents Marcia A. Bush; Xerox Palo Alto Research Center.....	376
SITE REPORTS	381

AUTHOR INDEX

Abney, S.233	Chevalier, H.191
Acero, A.69	Chinchor, N.56
Allan, J.392	Chipman, S.290
Allen, J.171, 420	Coccaro, N.211
Alleva, F.81	Cohen, P.370
Alshawi, H.217	Collier, R.341
Anastasakos, T.75	Cousseau, P.196
Anderson, A.25	Cowie, J.405
Aone, C.144	Croft, W.353, 418
Appelbaum, L.383	Dahl, D.19, 31
Appelt D.43,133	Danielson, D.413
Baker, James119, 191, 393	De Pijper, J.341
Baker, Janet119, 394	Della Pietra, S.202
Bamberg, P.191	Della Pietra, V.202
Bard, E.25	Dietzel, T.191
Bates, M.3, 19, 55, 387	Digalakis, V.87, 217, 415
Bear, J.43, 133	Doherty-Sneddon, G.25
Bernstein, J.412,413	Dolan, C.417
Biermann, A.177	Dowding, J.43
Blejer, H.144	Eisele, A.211
Bobrow, R.37	Elder, J.191
Bretan, I.217	Ferguson, G.171
Bridle, J.119	Fiscus, J.7
Brill, E.237	Fisher, W.7, 19
Broglio, J.353	Flank, S.144
Brown, P.202, 397	Frederking, R.196
Buckley, C.349, 392	Fussell, J.295
Bunker, R.303	Gallant, S.396
Bush, M.376	Gamback, B.217
Butzberger, J.87	Garofolo, J.7, 19
Caid, W.396	Gaston, J.370
Carlson, L.206	Gauvain, J.96
Carter, D.217	Gawron, J.43
Cherny, L.43	Gigley, H.287

Gillick, L.	119, 394	Leacock, C.	260, 303
Glass, J.	49	Lehnert, W.	417
Goldsmith, M.	202	Liddy, E.	159, 309, 358, 416
Goodine, D.	49	Linebarger, M.	31
Grannes, D.	196	Lippmann, R.	399
Grishman, R.	254, 299, 300, 407	Liu, F.	69
Guinn, C.	177	Lowe, S.	119
Hajic, J.	202	Lyberg, B.	217
Harman, D.	61	Macleod, C.	300
Hatzivassiloglou, V.	272	Makhoul, J.	75, 385, 387
Hindle, D.	243	Mandel, M.	191
Hipp, D.	177	Manganaro, L.	191
Hirschberg, J.	329	Marcus, M.	225, 419
Hirschman, L.	19, 401	Margolis, T.	191
Hobbs, J.	133, 157	Matsukawa, T.	227, 248
Hovy, E.	421	McCord, M.	127
Huang, X.	69, 81	McKee, D.	144
Hunnicke-Smith, D.	19, 412	McKenna, M.	159, 309
Hunt, M.	119	McKeown, K.	272, 391
Hwang, C.	138	McNair, A.	211
Hwang, M.	81	McVeary, K.	159, 358
Israel, D.	133	Mercer, R.	202
Ito, Y.	119	Miller, G.	303, 409
Jacobs, P.	395	Miller, S.	227
Joshi, A.	419	Mohanty, S.	202
Kaja, J.	217	Moore, J.	165
Kameyama, M.	133	Moore, R.	43, 125, 414
Kampmann, F.	191	Moran, D.	43
Karlgren, J.	217	Murveit, H.	87, 415
Kimball, O.	102	Myaeng, S.	416
Knight, K.	185	Nakatani, C.	329
Kubala, F.	75, 91	Newlands, A.	25
Kwok, K.	410	Nguyen, L.	75, 91
Lamel, L.	96	Nirenburg, S.	196
Lau, R.	108	Norton, L.	31
Lavie, A.	211	O'Connell, T.	206

Orloff, J.	119	Sotillo, C.	25
Ostendorf, M.	102, 315, 335, 388, 389	Sparck Jones, K.	347
Oviatt, S.	370	Stallard, D.	37
Paik, W.	159, 309, 358	Steedman, M.	419
Pallett, D.	7, 19, 402	Steele, E.	191
Passonneau, R.	391	Sterling, J.	254
Paul D.	67, 400	Stern, R.	69
Peskin, B.	119	Strzalkowski, T.	408
Phillips, M.	49	Sundheim, B.	56, 403
Placeway, P.	91	Tengi, R.	303
Polzin, T.	211	Thompson, H.	25
Price, P.	19, 217, 388, 414	Tomita, M.	211
Pulman S.	217	Tong, R.	383
Pustejovsky, J.	405	Towell, G.	260
Rayner, M.	217	Tsutsumi, J.	211
Reddy, R.	390	Tyson, M.	133
Resnik, P.	278	Tzoukermann, E.	19
Rogina, I.	211	Van Ess-Dykema, C.	285
Rohlicek, J.	384, 389	Van Santen, J.	323
Rose, C.	211	Veilleux, N.	335
Rosenfeld, R.	81,108	Vilain, M.	150
Roth, R.	119, 394	Voorhees, E.	260, 411
Roukos, S.	108, 398	Waibel, A.	183, 211
Rudnicky, A.	5, 19, 364	Waibel, N.	211
Sakai, S.	49	Wang, M.	370
Salton, G.	392	Ward, W.	211
Samuelsson, C.	217	Webber, B.	419
Sanderman, A.	341	Weinstein, C.	400
Scattone, F.	119	Weintraub, M.	87, 114, 415
Schubert, L.	138, 420	Weisedel R.	227, 363, 386
Schwartz, R.	75, 91, 385	White, J.	206
Seneff, S.	49	Wilks, Y.	404, 405, 406
Shinn, S.	144	Wolff, S.	300
Silverman, K.	317	Woszczyna, M.	211
Sloboda, T.	211	Yamron, J.	191, 393
Smith, R.	177	Yarowsky, D.	266

Yu, E.	159, 309, 358
Zavaliagos, G.	75
Zue, V.	49, 401