

**Fourth Conference
on
Applied
Natural Language
Processing**

**Association for
Computational Linguistics**

Proceedings of the Conference

**13-15 October 1994
Stuttgart, Germany**

Published by the Association for Computational Linguistics

©1994, Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Morgan Kaufmann Publishers
340 Pine Street, 6th Floor
San Francisco, CA 94104
1-800-745-7323

PREFACE

This volume represents the fourth proceedings of an “ACL Applied” conference. The contents of these pages illustrate some of the changes that have influenced us since the first such conference in 1983: a heightened emphasis on applications and commercial enterprise, integration of natural language processing with other technologies, and the emergence of innovative new applications and methods. The pace of change in the field has become so rapid that the community felt a need to have an applied conference on a regular basis; hence, this conference follows the third by only two years.

Never has the interplay between computational linguistics and applications been so active and harmonious. Emerging applications and even emerging commercial successes have driven research, for example, in multimedia systems, information retrieval, lexical representation, multi-lingual tools, and document processing. New methods in turn have led to improved applicability and commercial viability. The work reported here falls on both sides of an important boundary: the “technology push” that helps to create applications and markets, and the “market pull” that guides research and development, creates funding, and heightens the excitement in the field.

The process by which this proceedings has been created reflects some of the same changes that have driven the research. Much of the preparation, review, revision, and final transmission of the papers was done electronically, leading to the rapid, easy, and accurate exchange and reproduction of materials between points as distant as North America and Asia, as well as Eastern and Southern Europe. The same worldwide connectivity that makes this possible is also responsible for some of the advances in the field as well as for some emerging applications.

A number of people deserve special recognition for their contribution to this effort. Trisha Yannuzzi of the Institute for Research in Cognitive Science and Craig Reynolds of the GRASP Laboratory at the University of Pennsylvania helped to coordinate the program committee activities as well as the preparation of the proceedings, with help from Jodi Kerper and others at IRCS. Uwe Reyle, along with Christian Rohrer and others in Stuttgart, made the conference arrangements and did publicity as well as the cover layout. Fernando Pereira, Karen Sparck Jones, and others in the ACL provided logistical as well as moral support. These individuals contributed to a global, integrated project representative of the current state of the art.

Paul Jacobs, Program Chair
Philadelphia, Pennsylvania
September, 1994

SPONSORS

Association for Computational Linguistics

Gesellschaft für Informatik

IBM

LAPP-Kabel

Robert-Bosch-Stiftung

Landesgirokasse

Sun Microsystems

PROGRAM COMMITTEE

Susan Armstrong

Harry Bunt

Jim Cowie

Ido Dagan

Robert Ingria

Paul Jacobs, Chair

Richard Kittredge

Kazunori Muraki

Peter Norvig

Hans Joachim Novak

Martha Palmer

Manny Rayner

Donia Scott

Oliviero Stock

Annie Zaenen

ADDITIONAL REVIEWERS

Laurent Bourbeau

Rebecca Bruce

Fabio Ciravegna

Barbara Di Eugenio

Ted Dunning

David Farwell

Sebastian Goeser

Steve Helmreich

Wanying Jin

Jutta Kreys

Alberto Lavelli

Mark Liberman

Cecile Paris

Fabio Pianesi

Emanuele Pianta

Richard Power

Graham Russell

Carlo Strapparava

Jan Wiebe

Susuma Yasuda

TABLE OF CONTENTS

Text Generation

<i>Bilingual Generation of Job Descriptions from Quasi-Conceptual Forms</i> David E. Caldwell & Tatiana Korelsky	1
<i>Practical Issues in Automatic Documentation Generation</i> Kathleen McKeown, Karen Kukich & James Shaw	7

Document Image Understanding

<i>Language Determination: Natural Language Processing from Scanned Document Images</i> Penelope Sibun & A. Lawrence Spitz	15
<i>Modeling Content Identification from Document Images</i> Takehiro Nakayama	22

Machine Translation (Methods)

<i>Machine Translation of Sentences with Fixed Expressions</i> Naoto Katoh & Teruaki Aizawa	28
<i>Termight: Identifying and Translating Technical Terminology</i> Ido Dagan & Ken Church	34
<i>Symmetric Pattern Matching Analysis for English Coordinate Structures</i> Akitoshi Okumura & Kazunori Muraki	41

Tagging Models

<i>Tagging Accurately — Don't Guess If You Know</i> Pasi Tapanainen & Atro Voutilainen	47
<i>Does Baum-Welch Re-estimation Help Taggers?</i> David Elworthy	53
<i>Improving Language Models By Clustering Training Sentences</i> David Carter	59

Text Processing

<i>Exploiting Sophisticated Representations for Document Retrieval</i> Steven Finch	65
<i>Combination of Symbolic and Statistical Approaches for Grammatical Knowledge Acquisition</i> Masaki Kiyono & Jun'ichi Tsujii	72
<i>Adaptive Sentence Boundary Disambiguation</i> David D. Palmer & Marti A. Hearst	78
<i>Acquiring Knowledge from Encyclopedic Texts</i> Fernando Gomez, Richard Hull & Carlos Segami	84

Machine Translation (Systems)

<i>A Successful Case of Computer Aided Translation</i> Miguel Filgueiras	91
---	----

<i>Three Heads Are Better Than One</i>	
Robert Frederking & Sergei Nirenburg	95
<i>Real-Time Spoken Language Translation Using Associative Processors</i>	
Kozo Oi, Eiichiro Sumita, Osamu Furuse, Hitoshi Iida & Tetsuya Higuchi	101
Robust Parsing	
<i>Yet Another Chart-Based Technique for Parsing Ill-Formed Input</i>	
Tsuneaki Kato	107
<i>Recycling Terms into a Partial Parser</i>	
Christian Jacquemin	113
Interface Applications	
<i>Resolving Anaphora in a Portable Natural Language Front End to Databases</i>	
Flavia A. Barros & Anne Deroeck	119
<i>Upholding the Maxim of Relevance during Patient-Centered Activities</i>	
Abigail S. Gertner, Bonnie L. Webber & John R. Clarke	125
<i>The Delphi Natural Language Understanding System</i>	
Madeleine Bates, Robert Bobrow, Robert Ingria & David Stallard	132
<i>Understanding Location Descriptions in the Lei System</i>	
David N. Chin, Matthew McGranaghan & Tung-Tse Chen	138
Lexical Processing	
<i>Tagging and Morphological Disambiguation of Turkish Text</i>	
Kemal Oflazer & İlker Kuruöz	144
<i>A Robust Category Guesser for Dutch Medical Language</i>	
Peter Spyns	150
<i>Handling Japanese Homophone Errors in Revision Support System for Japanese Texts; Revise</i>	
Masahiro Oku	156
Text Categorization and Retrieval	
<i>A Probabilistic Model for Text Categorization: Based on a Single Random Variable with Multiple Values</i>	
Makoto Iwayama & Takenobu Tokunaga	162
<i>Robust Text Processing in Automated Information Retrieval</i>	
Tomek Strzalkowski	168
<i>Might a Semantic Lexicon Support Hypertextual Authoring?</i>	
Roberto Basili, Fabrizio Grisoli & Maria Teresa Pazienza	174
Posters	
<i>Improving Chinese Tokenization with Linguistic Filters on Statistical Lexical Acquisition</i>	
Dekai Wu & Pascale Fung	180
<i>Reference Resolution in Newspaper Articles</i>	
Takahiro Wakao	182

<i>Automatic Acquisition of Semantic Attributes for User Defined Words in Japanese to English Machine Translation</i>	
Satoru Ikehara, Satoshi Shirai, Akio Yokoo, Francis Bond & Yoshie Omi	184
<i>Degraded Text Recognition Using Word Collocation and Visual Inter-Word Constraints</i>	
Tao Hong & Jonathan J. Hull	186
<i>Using Syntactic Dependencies for Word Alignment</i>	
Fathi Debili, Elyès Sammouda & Adnane Zribi	188
<i>English Adverb Generation in Japanese to English Machine Translation</i>	
Kentaro Ogura, Francis Bond & Satoru Ikehara	190
<i>A Practical Evaluation of an Integrated Translation Tool During a Large Scale Localisation Project</i>	
Reinhard Schäler	192
<i>Spelling Correction in Agglutinative Languages</i>	
Kemal Oflazer & Cemalettin Güzey	194
<i>Integration of Example-Based Transfer and Rule-Based Generation</i>	
Susumu Akamine, Osamu Furuse & Hitoshi Iida	196
<i>An Evaluation of a Method to Detect and Correct Erroneous Characters in Japanese Input Through an OCR Using Markov Models</i>	
Tetsuo Araki, Satoru Ikehara, Nobuyuki Tsukahara & Yasunori Komatsu	198
<i>Multifunction Thesaurus for Russian Word Processing</i>	
Igor A. Bolshakov	200
<i>Representing Knowledge for Planning Multisentential Text</i>	
Jose Coch & Raphael David	203
<i>Guided Sentences Composition for Disabled People</i>	
Robert Pasero, Nathalie Richardet & Paul Sabatier	205
<i>An Interactive Rewriting Tool for Machine Acceptable Sentences</i>	
Hideki Hirakawa, Kouichi Nomura & Mariko Nakamura	207
<i>TECHDOC: Multilingual Generation of Online and Offline Instructional Text</i>	
Dietmar Rösner & Manfred Stede	209
<i>An Inheritance-Based Lexicon for Message Understanding Systems</i>	
Lynne J. Cahill	211
<i>Industrial Applications of Unification Morphology</i>	
Gábor Prózéký	213
<i>Sublanguage Engineering in the FoG System</i>	
Richard Kittredge, Eli Goldberg, Myunghee Kim & Alain Polguère	215

AUTHOR INDEX

Aizawa, Teruaki	28	Komatsu, Yasunori	198
Akamine, Susumu	196	Korelsky, Tatiana	1
Araki, Tetsuo	198	Kukich, Karen	7
Barros, Flavia A.	119	Kuruöz, İlker	144
Basili, Roberto	174	McGranaghan, Matthew	138
Bates, Madeleine	132	McKeown, Kathleen	7
Bobrow, Robert	132	Muraki, Kazunori	41
Bolshakov, Igor A.	200	Nakamura, Mariko	207
Bond, Francis	184, 190	Nakayama, Takehiro	22
Cahill, Lynne J.	211	Nirenburg, Sergei	95
Caldwell, David E.	1	Nomura, Kouichi	207
Carter, David	59	Oflazer, Kemal	144, 194
Chen, Tung-Tse	138	Ogura, Kentaro	190
Chin, David N.	138	Oi, Kozo	101
Church, Ken	34	Oku, Masahiro	156
Clarke, John R.	125	Okumura, Akitoshi	41
Coch, Jose	203	Omi, Yoshie	184
Dagan, Ido	34	Palmer, David D.	78
David, Raphael	203	Pasero, Robert	205
Debili, Fathi	188	Pazienza, Maria Teresa	174
DeRoeck, Anne	119	Polguère, Alain	215
Elworthy, David	53	Prószéky, Gábor	213
Filgueiras, Miguel	91	Richardet, Nathalie	205
Finch, Steven	65	Rösner, Dietmar	209
Frederking, Robert	95	Sabatier, Paul	205
Fung, Pascale	180	Sammouda, Elyès	188
Furuse, Osamu	101, 196	Schäler, Reinhard	192
Gertner, Abigail S.	125	Segami, Carlos	84
Goldberg, Eli	215	Shaw, James	7
Gomez, Fernando	84	Shirai, Satoshi	184
Grisoli, Fabrizio	174	Sibun, Penelope	15
Güzey, Cemaleddin	194	Spitz, A. Lawrence	15
Hearst, Marti A.	78	Spyns, Peter	150
Higuchi, Tetsuya	101	Stallard, David	132
Hirakawa, Hideki	207	Stede, Manfred	209
Hong, Tao	186	Strzalkowski, Tomek	168
Hull, Jonathan J.	186	Sumita, Eiichiro	101
Hull, Richard	84	Tapanainen, Pasi	47
Iida, Hitoshi	101, 196	Tokunaga, Takenobu	162
Ikehara, Satoru	184, 190, 198	Tsujii, Jun'ichi	72
Ingria, Robert	132	Tsukahara, Nobuyuki	198
Iwayama, Makoto	162	Voutilainen, Atro	47
Jacquemin, Christian	113	Wakao, Takahiro	182
Kato, Tsuneaki	107	Webber, Bonnie L.	125
Katoh, Naoto	28	Wu, Dekai	180
Kim, Myunghee	215	Yokoo, Akio	184
Kittredge, Richard	215	Zribi, Adnane	188
Kiyono, Masaki	72		