

*The 18th International Conference
on Computational Linguistics*



Proceedings of the Conference

*Universität des Saarlandes
Saarbrücken, Germany
31 July - 4 August 2000*

*Proceedings of
the 18th International Conference
on Computational Linguistics*



Proceedings of the Conference

Universität des Saarlandes

Saarbrücken, Germany

31 July - 4 August 2000

© 2000 DFKI GmbH

ISBN 1-55860-717-X
(2 volumes)

Current ACL members may order copies of this and other ACL-related proceedings from:

Association for Computational Linguistics (ACL)
75 Paterson Street, Suite 9
New Brunswick, NJ 08901, USA
Tel: +1-732-342-9100
Fax: +1-732-342-9339
acl@aclweb.org or rasmusse@cs.rutgers.edu

All other orders (from libraries, institutions, and other individuals) should be addressed to:

Morgan Kaufmann Publishers
340 Pine Street, 6th Floor
San Francisco, CA 94104 USA
Tel: +1-415-392-2665 (for individuals)
+1-619-699-6339 (for libraries & institutions)
Fax: +1-415-982-2665
orders@mkp.com
Or order on line at www.mkp.com

COLING 2000 Sponsors

The Organisation Committee of COLING 2000 wishes to thank the following sponsors for their generous support:

Our institutional sponsors:

- Universität des Saarlandes
- Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)
- Deutsche Forschungsgemeinschaft
- Ministerium für Bildung, Kultur und Wissenschaft des Saarlandes
- Loria
- Université Nancy 2
- Ministère de la Recherche Français
- Centre Universitaire de Luxembourg

Our corporate sponsors:

Main sponsors:

- Interprice
- XtraMind

Other supporters:

- DaimlerChrysler AG
- Google
- IDS Scheer AG
- Lexeme
- Microsoft Research
- Siemens AG
- Xerox PARC
- YY

Address of Welcome
by the Federal Minister of Education and Research,
Ms Edelgard Bulmahn, to the International Conference on Computational Linguistics –
Coling 2000

I am very happy that the International Conference on Computational Linguistics – Coling 2000 – is taking place in Saarbrücken, and I appreciate very much that the conference was prepared by three countries of the European Union, Germany, France and Luxembourg, and that some of its events are being held also at other places, namely Nancy and Luxembourg. After my recent opening of the German-French University in Saarbrücken, I consider this conference to be another important milestone on the way towards closer cooperation within a European region in the field of cutting-edge research.

Supporting speech and language technology has been a priority in the Federal Ministry of Education and Research since the mid-1980s. Research funds to the amount of about 165 million DM have so far been made available for this purpose, while for Artificial Intelligence as a whole about 650 million DM have been made available. The first funding projects of speech processing addressed simple spoken dialog and information systems. Since the beginning of the 1990s, speech and language technology research has been conducted on the basis of interdisciplinary cooperation between the fields of computational neuroscience. The emphasis is on the syntactic and semantic recognition and translation of spontaneous speech. Since 1993 research on speech and language technology in Germany has been concentrated in the lead project VERBMOBIL, in which 23 research teams from science and industry work together. The project is now being concluded and its results will be presented at this conference.

International experts confirm that, as a result of this research support, Germany is one of the leaders world-wide in the areas of speech-to-speech translation, robust semantic speech processing, information extraction and document analysis. A broad research infrastructure in this field is available for cooperation with industry. A large number of new start-up companies in the field of speech and language technology, research teams of foreign companies in Germany, the massive expansion of language technology groups in large German IT companies, and numerous new speech technology products document the commercial success of this support effort.

On the basis of the research results from speech and language technology in Germany, I have so far launched five major lead projects on human-computer interaction in the knowledge-based society, with a funding volume of 132 million DM. With these projects, new interaction possibilities between humans and IT systems via speech, gestures, facial expressions, and haptic signals are to be opened up.

I wish the International Conference on Computational Linguistics in Saarbrücken every success.



Edelgard Bulmahn

Preface: President of the ICCL and Programme Chair

This is the 18th International Conference on Computational Linguistics and, for myself and a few others, the 18th occasion on which I experience a growing sense of excitement at the prospect of meeting old friends, making new ones, and learning about the new discoveries and inventions that my colleagues have made in the last two years.

The late Hans Karlgren invented the name “Coling” as an obvious contraction of “computational linguistics”, and also in memory of the vagrant hero of a well-known Swedish comic strip who went by that name. The term “computational linguistics” itself was coined only a few years earlier by the late David Hays to refer to a field of endeavor whose creation had been recommended by the Automatic Language Processing Advisory Board (ALPAC) to provide a more solid theoretical foundation for work on machine translation. During the five or so years that intervened, the International Committee on Computational Linguistics stumbled into existence and organised its first conference in New York in 1965.

During the 35 years since that first meeting, our field has become broader and deeper, and our students and colleagues can now command good salaries throughout much of the world. For evidence of the vigor of the field, you have only to look around you. The day before the opening of this meeting, a ceremony in Saarbrücken will mark the termination of the 8-year, 89-million dollar Verbmobil project on speech-to-speech translation that involved at one time or another, some 900 people. It will therefore be no surprise that Germany is one of the three countries to whom we owe such a debt of gratitude for the organisation of the meeting.

The programme committee decided to make some modest innovations of its own this year, determining for the first time to conduct its business entirely through the Internet. Electronic submission of papers was strongly encouraged, against the advice of several experienced and cautious people who predicted chaos for the conference and a mental breakdown for me. There were, of course, problems. But things went more smoothly than anyone expected and we will know how to do things better next time. No submission was sent for review as hard copy, the handful that arrived on paper being scanned to make the digital version that reviewers then accessed through the world-wide web.

323 regular papers, and about 100 project notes and demonstration proposals were received from 22 countries. Each was read by three members of one of the 11 panels of reviewers, totalling 222 people altogether. My gratitude to those dedicated and able people knows no bounds. 110 regular papers, 24 project notes, and 10 demonstration proposals were accepted, or about a third of the submissions. As one who has read many of the papers, I can assure you that the quality of the presentations is likely to be even higher than even this good ratio suggests. So I hope you share some of my excitement at what will be either the last Coling of the old millenium or the first of the new one, but surely one of a series that has only just begun.

Martin Kay

Preface : Chair of the Organisation Committee

Every COLING conference I have attended has been special in its own way. COLING 2000 is special in three respects: time, place and people.

For a number of years, we have learned to live with predictions that our discipline was just about to enter a completely new era, a prosperous period of commercialisation. But not until now has it finally started happening on a broad scale. COLING 2000 seems to be the COLING of transition. That this transition occurs simultaneously with the start of a new millenium is a coincidence whose mnemonic value might only be appreciated by the students of our students. It is an exciting time. On one hand, we are still far away from unravelling the miracle of human language, but on the other hand, our results are already becoming useful to a broad customer base. This has led to the creation of numerous new companies in language technology. Quite a few of these start-ups are represented at our conference. In these proceedings some well-known names appear with new affiliations. Conference name tags link familiar faces with the imaginative names of small but ambitious corporations.

We could sense the dynamics of commercialisation when we contacted companies to solicit support for our conference. Although all industrial colleagues we contacted on a personal basis were extremely sympathetic to our good cause and expressed great interest in COLING, some companies were simply so busy coping with growth and economic success that they could not turn their interest into active support and divert resources at this very time of rapid development. We are all the more grateful to the businesses that were able to help us. As these proceedings go to press, we have received confirmation of support from the sponsors acknowledged on page i.

The second factor which makes COLING 2000 special is the venue of the event in the very heart of Europe. It is the first COLING to take place in three neighbouring countries. Nowadays you have to look very hard to detect the exact location of the political borders when you move from one conference site to the other, yet for nearly one thousand years these borders were the object of bloody wars. Nobody can count the human lives that were lost in the repeated silly attempts to move the borders by a few miles in one direction or the other. If you should have the possibility to go hiking in the area, you will see the remnants of fortresses, trenches, bunkers and graves. Only for the last forty-three years have the borders remained stable while the inhabitants have been enjoying the longest period of peace in the past millenium.

It was a great pleasure for us to work with our Programme Chair Martin Kay. The collaboration among the three sites went very smoothly. Our main partners were Laurent Romary in Nancy and Antonio Sanfilippo in Luxembourg. The main workload in the organisation was carried by Michelle Carnell, Carmen Görl, Corinna Johanns, Jasmin Schneider and Jan Timm. They were supported by Uta Merkle and Marion Metzen of the University's Contact Office for Knowledge and Technology Transfer. I cannot list here all the students, researchers and faculty who helped us with the preparations. For the computational linguists of Saarbrücken, Nancy and Luxembourg it is an honour and a great pleasure to welcome the international community to the SaarLorLux region. On behalf of my local and regional colleagues I wish you a successful and pleasant conference.

Hans Uszkoreit

Programme Committee and Reviewers by Subject Area

Programme Chair: Martin Kay

Phonology and morphology: Kemal Oflazer (chair), Jan Amtrup, Ken Beesley, Steven Bird, Key-Sun Choi, Walter Daelemans, Tomaz Erjavec, Dafydd Gibbon, Toru Hisamitsu, Arvi Hurskainen, Sharon Inkelas, Lauri Karttunen, George Kiraz, Ewan Klein, Kimmo Koskenniemi, Jong-hyeok Lee, Yuji Matsumoto, Andrei Mikheev, Marge McShane, Gabor Proszeky, Svetlana Sheremetyeva, Markus Walther, Dekai Wu, Kenji Yoshimura

Discourse and dialogue: Bonnie Webber (chair), Claire Gardent, Andrew Kehler, Alistair Knott, Alex Lascarides, Johanna Moore, Massimo Poesio, Laurent Romary, Claudia Soria, Michael Strube, Sandra Carberry, Jennifer Chu-Carroll, Barbara Di Eugenio, Susan Haller, Jan van Kuppevelt, Susan McRoy, Akira Shimazu, Matthew Stone, Rich Thomason.

Information extraction: Winfried Lenders (chair), Douglas E. Appelt, Sophia Ananiadou, Key-Sun Choi, Fabio Ciravegna, Fumiyo Fukumoto, Ulricke Hahn, Karin Harbusch, Eduard H. Hovy, Chu-Ren Huang, Gerhard Knorz, Rainer Kühlen, H. D. Luckhardt, John Matiassek, Günter Neumann, Dietmar Roesner, Manfred Stede, Harald Trost, Harald H. Zimmermann.

Syntactic Issues and Formalisms: Chris Manning (chair), Geert-Jan Kruijff, Dekang Lin, Srinivas Bangalore, Chris Brew, Ted Briscoe, Mark Dras, Dominique Estival, Takao Gunji, Michael Johnston, Andreas Kathol, Guido Minnen, Mick O'Donnell, Ed Stabler, Miriam Butt, Detmar Meurers, Anette Frank.

Parsing and Generation: John Carroll (chair), Steve Abney, Roberto Basili, Gosse Bouma, Lynne Cahill, Ann Copestake, Robert Dale, Dan Flickinger, Mark Johnson, Mark Hepple, Alon Lavie, Rob Malouf, Suresh Manandhar, Guido Minnen, Bob Moore, Mark-Jan Nederhof, Anton Nijholt, Stephan Oepen, Gerald Penn, Ehud Reiter, Giorgio Satta, Mark Steedman, Michael Zock, Chris Brew, Kentaro Inui, Miles Osborne, Srinivas Bangalore, Marc Light.

Semantics and Pragmatism: Steve Pullman (chair), Patrick Blackburn, Robin Cook, Dick Crouch, Claire Gardent, Jerry Hobbs, Ewan Klein, Shalom Lappin, Ian Lewin, David Milward, Uwe Reyle, Ann Copestake, Manfred Pinkal, Johan Bos, Kees van Deemter, Reinhard Muskens, Bonnie Webber, Len Schubert, Jan van Eijck, Jan Jaspars, Graeme Ritchie.

Computational Paradigms: Jun-ichi Tsujii, (chair), David Carter, Key-Sun Choi, G.B. Lee, John Carroll, K. T. Lua, Satoshi Sekine, Keh-Yih Su, Chu-Ren Huang, Hadar Shemtov, K. Torisawa, Y. Matsumoto, H. Isahara, K. Shirai, S. Kurohashi, Y. Nakao, N. Collier, H. Nakagawa.

Machine Translation: Sergei Nirenburg (chair), Elliott Macklovitch, Rod Johnson, Francis Bond, Harold Somers, Mori Rimon, Michael Blekhan, Doug Arnold, Yorick Wilks, David Farwell, Boyan Onyshkevych, Lynn Carlson, Laurie Gerber, Doug Jones, Dan Tufis, Arturo Trujillo, Graham Russell, Maghi King, Fred Popowich, Timothy Baldwin, Pierrette Bouillon, Steve Helmreich, Pierre Isabelle, Martha Palmer, Leonid Iomdin, Davide Turcato, Eric Gaussier.

Multilingual: Christian Boitet (chair), Malek Boualem, Jean-Pierre Chanod, Fathi Debili, Christian Fluhr, Gregory Grefenstette, Mathieu Lafourcade, Kazunori Muraki, Antoine Ogonowski, Emmanuel Planas, N. Puntikov, Jörg Schütz, Mark Seligman, Gilles Sérasset, Tong Loong Cheong, Jean Veronis, Rémi Zajac.

Programming Tools: Gertjan van Noord (chair), John Bateman, Bill Black, Thorsten Brants, Jo Calder, David Carter, Martin Cooke, Rob Gaizauskas, Dale Gerdemann, Thilo Götze, Edmund Grimley Evans, Kôiti Hasida, Ronald Kaplan, Ulli Krieger, Torbjorn Lager, Alberto Lavelli, Mehryar Mohri, Günter Neumann, Stephan Oepen, Gerald Penn, Mike Rossner, Anoop Sarkar, Hadar Shemtov, Neil Simpkins, Shuli Wintner, Rémi Zajac, Phil Resnik.

Corpus: Susan Armstrong (chair), Thorsten Brants, Chris Brew, Jean-Pierre Chanod, Ken Church, Andreas Eisele, Eva Ejerhed, Benoit Haber, David Hull, Nancy Die, Christian Jacquemin, Martin Jansche, Maria Lapata, André Kempe, Tony McEnery, Paola Merlo, Hermann Ney, Laurent Romary, Christer Samuelsson, Ludovic Tanguy, Henry Thompson, Pim van der Eijk, Atro Voutilainen, David Yarowsky, Rémi Zajac, Steve Abney.

Other members of the Programme Committee: Eva Hajičová

Table of Contents

COLING 2000 Sponsors	I
Welcome Address	III
Preface: President of the ICCL and Programme Chair	IV
Preface: Chair of the Organisation Committee	V
Programme Committee and Reviewers by Subject Area	VI
Table of Contents	VII
Regular Papers	VIII
Project Notes and Demos	XIV
Author Index	XVI

Regular Papers

I. Aduriz, E. Agirre, I. Aldezabal, I. Alegria, X. Arregi, J. M. Arriola, X. Artola, K. Gojenola, A. Maritxalar, K. Sarasola, M. Urkia <i>A Word-grammar based Morphological Analyzer for Agglutinative Languages</i>	1
Paolo Allegrini, Simonetta Montemagni, Vito Pirrelli <i>Learning Word Clusters from Data Types</i>	8
Ion Androutsopoulos, Robert Dale <i>Selectional Restrictions in HPSG</i>	15
Masayuki Asahara, Yuji Matsumoto <i>Extended Models and Tools for High-performance Part-of-speech Tagger</i>	21
Tania Avgustinova, Hans Uszkoreit <i>An Ontology of Systematic Relations for a Shared Grammar of Slavic</i>	28
Timothy Baldwin, Hozumi Tanaka <i>The Effects of Word Order and Segmentation on Translation Retrieval Performance</i>	35
Srinivas Bangalore, Owen Rambow <i>Exploiting a Probabilistic Hierarchical Model for Generation</i>	42
Petra Barg, James Kilbury <i>Incremental Identification of Inflectional Types</i>	49
José-Miguel Benedí, Joan-Andreu Sánchez <i>Combination of N-Grams and Stochastic Context-Free Grammars for Language Modeling</i>	55
Rens Bod <i>An Empirical Evaluation of LFG-DOP</i>	62
Rens Bod <i>Parsing with the Shortest Derivation</i>	69
Branimir K. Boguraev, Mary S. Neff <i>The Effects of Analysing Cohesion on Document Summarisation</i>	76
Igor Boguslavsky, Nadezhda Frid, Leonid Iomdin, Leonid Kreidlin, Irina Sagalova, Victor Sizov <i>Creating a Universal Networking Language Module within an Advanced NLP System</i>	83
Francis Bond, Kyonghee Paik <i>Reusing an Ontology to Generate Numeral Classifiers</i>	90
Lars Borin <i>You'll Take the High Road and I'll Take the Low Road: Using a Third Language to Improve Bilingual Word Alignment</i>	97
António Branco <i>Binding Constraints as Instructions of Binding Machines</i>	104
Thorsten Brants, Matthew Crocker <i>Probabilistic Parsing and Psychological Plausibility</i>	111
Norbert Bröker <i>The Use of Instrumentation in Grammar Engineering</i>	118
Ralf D. Brown <i>Automated Generalization of Translation Examples</i>	125
Caroline Brun <i>A Client/Server Architecture for Word Sense Disambiguation</i>	132
Eva Buráňová, Eva Hajičová, Petr Sgall <i>Tagging of Very Large Corpora: Topic-Focus Articulation</i>	139
Farid Cerbah <i>Exogeneous and Endogeneous Approaches to Semantic Categorization of Unknown Technical Terms</i>	145

Gerald Chao, Michael G. Dyer <i>Word Sense Disambiguation of Adjectives Using Probabilistic Networks</i>	152
Hsin-Hsi Chen, Chuan-Jie Lin <i>A Multilingual News Summarizer</i>	159
Hsin-Hsi Chen, Shih-Chung Tsai, Jin-He Tsai <i>Mining Tables from Large Scale HTML Texts</i>	166
Keh-Jiann Chen, Chao-jan Chen <i>Automatic Semantic Classification for Chinese Unknown Compound Nouns</i>	173
Kenneth W. Church <i>Empirical Estimates of Adaptation: The Chance of Two Noriegas is Closer to $p/2$ than p^2</i>	180
Massimiliano Ciaramita, Mark Johnson <i>Explaining Away Ambiguity: Learning Verb Selectional Preference with Bayesian Networks</i>	187
Stephen Clark, David Weir <i>A Class-based Probabilistic Approach to Structural Disambiguation</i>	194
Nigel Collier, Chikashi Nobata, Jun-ichi Tsujii <i>Extracting the Names of Genes and Gene Products with a Hidden Markov Model</i>	201
Dan Cristea, Nancy Ide, Daniel Marcu, Valentin Tablan <i>An Empirical Investigation of the Relation Between Discourse Structure and Co-Reference</i>	208
Béatrice Daille <i>Morphological Rule Induction for Terminology Acquisition</i>	215
Kees van Deemter, Richard Power <i>Authoring Multimedia Documents Using WYSIWYM Editing</i>	222
Hervé Déjean <i>Theory Refinement and Natural Language Learning</i>	229
Guy De Pauw <i>Aspects of Pattern-matching in Data-Oriented Parsing</i>	236
Marc Dymetman, Veronika Lux, Aarne Ranta <i>XML and Multilingual Document Authoring: Convergent Trends</i>	243
Christian Ebert, Marcus Kracht <i>Formal Syntax and Semantics of Case Stacking Languages</i>	250
Jason Eisner <i>Directional Constraint Evaluation in Optimality Theory</i>	257
Cécile Fabre, Christian Jacquemin <i>Boosting Variant Recognition with Light Semantics</i>	264
Udo Hahn, Martin Romacker <i>An Integrated Model of Semantic and Conceptual Interpretation from Dependency Structures</i>	271
Eva Hajičová, Markéta Ceplová <i>Deletions and Their Reconstruction in Tectogrammatical Syntactic Tagging of Very Large Corpora</i>	278
Dilek Z. Hakkani-Tür, Kemal Oflazer, Gökhan Tür <i>Statistical Morphological Disambiguation for Agglutinative Languages</i>	285
Sanda M. Harabagiu, Marius A. Paşca, Steven J. Maiorano <i>Experiments with Open-Domain Textual Question Answering</i>	292
Vasileios Hatzivassiloglou, Janyce M. Wiebe <i>Effects of Adjective Orientation and Gradability on Sentence Subjectivity</i>	299
Renate Henschel, Hua Cheng, Massimo Poesio <i>Pronominalization Revisited</i>	306

Hideki Hirakawa, Kenji Ono, Yumiko Yoshimura <i>Automatic Refinement of a POS Tagger Using a Reliable Parser and Plain Text Corpora</i>	313
Toru Hisamitsu, Yoshiki Niwa, Jun-ichi Tsujii <i>A Method of Measuring Term Representativeness - Baseline Method Using Co-occurrence Distribution -</i>	320
Véronique Hoste, Steven Gillis, Walter Daelemans <i>A Rule Induction Approach to Modeling Regional Pronunciation Variation</i>	327
Matthew Hurst, Tetsuya Nasukawa <i>Layout and Language: Integrating Spatial and Linguistic Knowledge for Layout Understanding Tasks</i>	334
Yumi Ichimura, Yoshimi Saito, Kazuhiro Kimura, Hideki Hirakawa <i>Kana-Kanji Conversion System with Input Support Based on Prediction</i>	341
Takashi Inui, Kentaro Inui <i>Committee-based Decision Making in Probabilistic Partial Parsing</i>	348
Mark Johnson, Brian Roark <i>Compact Non-left-recursive Grammars Using the Selective Left-corner Transform and Factoring</i>	355
Michael Johnston <i>Deixis and Conjunction in Multimodal Systems</i>	362
Michael Johnston, Srinivas Bangalore <i>Finite-state Multimodal Parsing and Understanding</i>	369
Douglas A. Jones, Gregory M. Rusk <i>Toward a Scoring Function for Quality-Driven Machine Translation</i>	376
SungYoung Jung, SungLim Hong, Eunok Paek <i>An English to Korean Transliteration Model of Extended Markov Window</i>	383
Gary Kacmarcik, Chris Brockett, Hisami Suzuki <i>Robust Segmentation of Japanese Text into a Lattice for Parsing</i>	390
Kyo Kageura, Keita Tsuji, Akiko N. Aizawa <i>Automatic Thesaurus Generation through Multiple Filtering</i>	397
Hiroyuki Kaji, Yasutsugu Morimoto, Toshiko Aizono, Noriyuki Yamasaki <i>Corpus-dependent Association Thesauri for Information Retrieval</i>	404
Hiroshi Kanayama, Kentaro Torisawa, Yutaka Mitsuishi, Jun-ichi Tsujii <i>A Hybrid Japanese Parser with Hand-crafted Grammar and Statistics</i>	411
In-Ho Kang, GilChang Kim <i>English-to-Korean Transliteration using Multiple Unbounded Overlapping Phoneme Chunks</i>	418
Ronald M. Kaplan, Jürgen Wedekind <i>LFG Generation Produces Context-Free Languages</i>	425
Daisuke Kawahara, Nobuhiro Kaji, Sadao Kurohashi <i>Japanese Case Structure Analysis by Unsupervised Construction of a Case Frame Dictionary</i>	432
Seonho Kim, Juntae Yoon, Mansuk Song <i>Structural Feature Selection for English-Korean Statistical Machine Translation</i>	439
Alexandra Kinyon <i>Hypertags</i>	446
Youngjoong Ko, Jungyun Seo <i>Automatic Text Categorization by Unsupervised Learning</i>	453
Alexander Koller, Joachim Niehren <i>On Underspecified Processing of Dynamic Semantics</i>	460
Kazunori Komatani, Tatsuya Kawahara <i>Flexible Mixed-Initiative Dialogue Management using Concept-Level Confidence Measures of Speech Recognizer Output</i>	467

Geert-Jan Kruijff, Elke Teich, John Bateman, Ivana Kruijff-Korbayová, Hana Skoumalová, Serge Sharoff, Lena Sokolova, Tony Hartley, Kamenka Staykova, Jiří Hana <i>Multilinguality in a Text Generation System for Three Slavic Languages</i>	474
Sang-Zoo Lee, Jun-ichi Tsujii, Hae-Chang Rim <i>Lexicalized Hidden Markov Models for Part-of-Speech Tagging</i>	481
Yves Lepage <i>Languages of Analogical Strings</i>	488
Chin-Yew Lin, Eduard Hovy <i>The Automated Acquisition of Topic Signatures for Text Summarization</i>	495
Diane J. Litman, Michael S. Kearns, Satinder Singh, Marilyn A. Walker <i>Automatic Optimization of Dialogue Management</i>	502
Qing Ma, Masaki Murata, Kiyotaka Uchimoto, Hitoshi Isahara <i>Hybrid Neuro and Rule-Based Part of Speech Taggers</i>	509
Christos Malavazos, Stelios Piperidis <i>Application of Analogical Modelling to Example Based Machine Translation</i>	516
Daniel Marcu <i>Extending a Formal and Computational Model of Rhetorical Structure Theory with Intentional Structures à la Grosz and Sidner</i>	523
Diana Maynard, Sophia Ananiadou <i>Identifying Terms by their Family and Friends</i>	530
Adam Meyers, Michiko Kosaka, Ralph Grishman <i>Chart-Based Transfer Rule Application in Machine Translation</i>	537
Kyonghi Moon, Jong-Hyeok Lee <i>Representation and Recognition Method for Multi-Word Translation Units in a Korean-to-Japanese MT System</i>	544
Frank Morawietz <i>Chart Parsing and Constraint Programming</i>	551
Shinsuke Mori, Masafumi Nishimura, Nobuyasu Itoh, Shiho Ogino, Hideo Watanabe <i>A Stochastic Parser Based on a Structural Word Prediction Model</i>	558
Masaki Murata, Kiyotaka Uchimoto, Qing Ma, Hitoshi Isahara <i>Bunsetsu Identification Using Category-Exclusive Rules</i>	565
Yukiko I. Nakano, Kenji Imamura, Hisashi Ohara <i>Taking Account of the User's View in 3D Multimodal Instruction Dialogue</i>	572
Shiho Nobesawa, Hiroaki Saito, Masakazu Nakanishi <i>Automatic Semantic Sequence Extraction from Unrestricted Non-Tagged Texts</i>	579
Miles Osborne <i>Estimation of Stochastic Attribute-Value Grammars using an Informative Sample</i>	586
Volume 2	
Jong C. Park, Hyung Joon Cho <i>Informed Parsing for Coordination with Combinatory Categorical Grammar</i>	593
Guy Perrier <i>Interaction Grammars</i>	600
Paul Piwek <i>A Formal Semantics for Generating and Editing Plurals</i>	607
Ferran Pla, Antonio Molina, Natividad Prieto <i>Tagging and Chunking with Bigrams</i>	614

Emmanuel Planas, Osamu Furuse <i>Multi-level Similar Segment Matching Algorithm for Translation Memories and Example-Based Machine Translation</i>	621
Sylvain Pogodalla <i>Generation, Lambek Calculus, Montague's Semantics and Semantic Proof Nets</i>	628
Arjen Poutsma <i>Data-Oriented Translation</i>	635
Richard Power <i>Planning Texts by Constraint Satisfaction</i>	642
Detlef Prescher, Stefan Riezler, Mats Rooth <i>Using a Probabilistic Class-Based Lexicon for Lexical Ambiguity Resolution</i>	649
Julien Quint <i>A Formalism for Universal Segmentation of Text</i>	656
Allan Ramsay, Helen Seville <i>Unscrambling English word order</i>	663
Manny Rayner, Beth Ann Hockey, Frankie James, Elizabeth Owen Bratt, Sharon Goldwater, Jean Mark Gawron <i>Compiling Language Models from a Linguistically Motivated Unification Grammar</i>	670
Tobias Ruland <i>A Context-Sensitive Model for Probabilistic LR Parsing of Spoken Language with Transformation-Based Postprocessing</i>	677
Christer Samuelsson <i>A Statistical Theory of Dependency Syntax</i>	684
Anoop Sarkar, Daniel Zeman <i>Automatic Extraction of Subcategorization Frames for Czech</i>	691
Yutaka Sasaki, Yoshihiro Matsuo <i>Learning Semantic-Level Information Extraction Rules by Type-Oriented ILP</i>	698
Manabu Sassano, Takehito Utsuro <i>Named Entity Chunking Techniques in Supervised Learning for Japanese Named Entity Recognition</i>	705
Michael Schiehlen <i>Granularity Effects in Tense Translation</i>	712
Barry Schiffman, Kathleen R. McKeown <i>Experiments in Automated Lexicon Building for Text Searching</i>	719
Helmut Schmid, Sabine Schulte im Walde <i>Robust German Noun Chunking With a Probabilistic Context-Free Grammar</i>	726
Karl-Michael Schneider <i>Parsing Schemata for Grammars with Variable Number and Order of Constituents</i>	733
Leah Schroeder, Sandra Carberry <i>Realizing Expressions of Doubt in Collaborative Dialogue</i>	740
Sabine Schulte im Walde <i>Clustering Verbs Semantically According to their Alternation Behaviour</i>	747
Satoshi Sekine, Kiyotaka Uchimoto, Hitoshi Isahara <i>Backward Beam Search Algorithm for Dependency Analysis of Japanese</i>	754
Satoshi Sekine <i>Japanese Dependency Analysis using a Deterministic Finite State Transducer</i>	761
Gilles Sérasset, Christian Boitet <i>On UNL as the Future "html of the Linguistic Content" and the Reuse of Existing NLP Components in UNL-related Applications with the Example of a UNL-French Deconverter</i>	768

Helen Seville, Allan Ramsay <i>Making Sense of Reference to the Unfamiliar</i>	775
Sayori Shimohata <i>An Empirical Method for Identifying and Translating Technical Terminology</i>	782
Elvira I. Sicilia-Garcia, Ji Ming, F.J. Smith <i>A Dynamic Language Model Based on Individual Word Domains</i>	789
Gerardo Sierra, John McNaught <i>Extracting Semantic Clusters from the Alignment of Definitions</i>	795
Virach Sornlertlamvanich, Tanapong Potipiti, Thatsanee Charoenporn <i>Automatic Corpus-Based Thai Word Extraction with the C4.5 Learning Algorithm</i>	802
Efstathios Stamatatos, Nikos Fakotakis, George Kokkinakis <i>Text Genre Detection Using Common Word Frequencies</i>	808
Suzanne Stevenson, Paola Merlo <i>Automatic Lexical Acquisition Based on Statistical Distributions</i>	815
Hisami Suzuki, Chris Brockett, Gary Kacmarcik <i>Using a Broad-Coverage Parser for Word-Breaking in Japanese</i>	822
Elke Teich, Catherine I. Watson, Cécile Pereira <i>Matching a Tone-based and Tune-based Approach to English Intonation for Concept-to-Speech Generation</i>	829
Aristomenis Thanopoulos, Nikos Fakotakis, George Kokkinakis <i>Automatic Extraction of Semantic Relations from Specialized Corpora</i>	836
Dan Tidhar, Uwe Kuessner <i>Learning to Select a Good Translation</i>	843
Christoph Tillmann, Hermann Ney <i>Word Re-ordering and DP-based Search in Statistical Machine Translation</i>	850
Erik F. Tjong Kim Sang, Walter Daelemans, Hervé Déjean, Rob Koeling, Yuval Krymolowski, Vasin Punyakanok, Dan Roth <i>Applying System Combination to Base Noun Phrase Identification</i>	857
Shu-Chuan Tseng <i>Modelling Speech Repairs in German and Mandarin Chinese Spoken Dialogues</i>	864
Kiyotaka Uchimoto, Masaki Murata, Qing Ma, Satoshi Sekine, Hitoshi Isahara <i>Word Order Acquisition from Corpora</i>	871
Yoshihiro Ueda, Mamiko Oka, Takahiro Koyama, Tadanobu Miyauchi <i>Toward the "At-a-glance" Summary: Phrase-representation Summarization Method</i>	878
Masao Utiyama, Masaki Murata, Hitoshi Isahara <i>A Statistical Approach to the Processing of Metonymy</i>	885
Masao Utiyama, Kôiti Hasida <i>Multi-Topic Multi-Document Summarization</i>	892
Renata Vieira, Massimo Poesio <i>Corpus-Based Development and Evaluation of a System for Processing Definite Descriptions</i>	899
Hideo Watanabe, Sadao Kurohashi, Eiji Aramaki <i>Finding Structural Correspondences from Bilingual Parsed Corpus for Corpus-based Translation</i>	906
Hideo Watanabe <i>A Method for Accelerating CFG-Parsing by Using Dependency Information</i>	913
Maria Wolters, Donna K. Byron <i>Prosody and the Resolution of Pronominal Anaphora</i>	919
Kiyoshi Yamabana, Shinichi Ando, Kiyomi Mimura <i>Lexicalized Tree Automata-based Grammars for Translating Conversational Texts</i>	926

Kaoru Yamamoto, Yuji Matsumoto <i>Acquisition of Phrase-level Bilingual Correspondence using Dependency Structure</i>	933
Roman Yangarber, Ralph Grishman, Pasi Tapanainen, Silja Huttunen <i>Automatic Acquisition of Domain Knowledge for Information Extraction</i>	940
Alexander Yeh <i>More Accurate Tests for the Statistical Significance of Result Differences</i>	947
Juntae Yoon, Yoonkwan Kim, Mansuk Song <i>Identifying Temporal Expression and its Syntactic Role Using FST and Lexical Data from Corpus</i>	954
Menno van Zaanen <i>ABL: Alignment-Based Learning</i>	961
Klaus Zechner, Alex Waibel <i>DiaSumm: Flexible Summarization of Spontaneous Dialogues in Unrestricted Domains</i>	968
Qiang Zhou <i>Local Context Templates for Chinese Constituent Boundary Prediction</i>	975

Project Notes and Demos

Jan W. Amtrup, Karine Megerdoomian, Rémi Zajac <i>Rapid Development of Translation Tools: Application to Persian and Turkish</i>	982
Igor Boguslavsky, Svetlana Grigorieva, Nikolai Grigoriev, Leonid Kreidlin, Nadezhda Frid <i>Dependency Treebank for Russian: Concept, Tools, Types of Information</i>	987
Jo Calder <i>Thistle and Interarbora</i>	992
Michael Carl <i>A Model of Competence for Corpus-Based Machine Translation</i>	997
Jeongwon Cha, Geunbae Lee <i>Structural Disambiguation of Morpho-syntactic Categorical Parsing for Korean</i>	1002
Jim Cowie, Yevgeny Ludovik, Hugo Molina-Salgado, Sergei Nirenburg <i>The Week at a Glance - Cross-language Cross-document Information Extraction and Translation</i>	1007
Hoa Trang Dang, Karin Kipper, Martha Palmer <i>Integrating Compositional Semantics into a Verb Lexicon</i>	1011
Marc Dymetman, Frédéric Tendeau <i>Context-Free Grammar Rewriting and the Transfer of Packed Linguistic Representations</i>	1016
John Elliott, Eric Atwell, Bill Whyte <i>Language Identification in Unknown Signals</i>	1021
Kilian Foth, Wolfgang Menzel, Horia F. Pop, Ingo Schröder <i>An Experiment on Incremental Analysis Using Robust Parsing Techniques</i>	1026
Alexander Franz, Keiko Horiguchi, Lei Duan, Doris Ecker, Eugene Koontz, Kazami Uchida <i>An Integrated Architecture for Example-Based Machine Translation</i>	1031
Antje Helfrich, Bradley Music <i>Design and Evaluation of Grammar Checkers in Multiple Languages</i>	1036
Christopher Hogan, Robert Frederking <i>WebDIPLOMAT: A Web-Based Interactive Machine Translation System</i>	1041
Bernd Kiefer, Hans-Ulrich Krieger, Melanie Siegel <i>An HPSG-to-CFG Approximation of Japanese</i>	1046

Byeongchang Kim, Geunbae Lee <i>Decision-Tree Based Error Correction for Statistical Phrase Break Prediction in Korean</i>	1051
Esther König, Wolfgang Lezius <i>A Description Language for Syntactically Annotated Corpora</i>	1056
Yevgeny Ludovik, Ron Zacharski <i>MT and Topic-Based Techniques to Enhance Speech Recognition Systems for Professional Translators</i>	1061
Hiroshi Masuichi, Raymond Flournoy, Stefan Kaufmann, Stanley Peters <i>A Bootstrapping Method for Extracting Bilingual Text Pairs</i>	1066
Hidetsugu Nanba, Manabu Okumura <i>Producing More Readable Extracts by Revising Them</i>	1071
Rani Nelken, Nissim Francez <i>Querying Temporal Databases Using Controlled Natural Language</i>	1076
Sonja Nießen, Hermann Ney <i>Improving SMT Quality with Morpho-syntactic Analysis</i>	1081
Franz Josef Och, Hermann Ney <i>A Comparison of Alignment Models for Statistical Machine Translation</i>	1086
Matthias Rehm, Karl Ulrich Goecke <i>Perception, Concepts and Language: RoAD and IPaGe</i>	1091
Won-Ho Ryu, Jin-Dong Kim, Hae-Chang Rim, Heui-Seok Lim <i>KCAT: A Korean Corpus Annotating Tool Minimizing Human Intervention</i>	1096
Michael Schiehlen <i>Robust Semantic Construction</i>	1101
Satoshi Sekine, Yoshio Eriguchi <i>Japanese Named Entity Extraction Evaluation - Analysis of Results -</i>	1106
Svetlana Sheremetyeva, Sergei Nirenburg <i>Acquisition of a Language Computational Model for NLP</i>	1111
Jörg Spilker, Martin Klarner, Günther Görz <i>Processing Self Corrections in a Speech to Speech System</i>	1116
Benjamin K. T'sou, K. K. Sin, Samuel W. K. Chan, Tom B. Y. Lai, C. Lun, K. T. Ko, Gary K. K. Chan, Lawrence Y. L. Cheung <i>Jurilinguistic Engineering in Cantonese Chinese: An N-gram-based Speech to Text Transcription System</i>	1121
Carole Tiberius, Lynne Cahill <i>Incorporating Metaphonemes in a Multilingual Lexicon</i>	1126
Stephan Vogel, Hermann Ney <i>Construction of a Hierarchical Translation Memory</i>	1131
Derek Walker, Dominique Petitpierre, Susan Armstrong <i>XMLTrans: a Java-based XML Transformation Language for Structured Data</i>	1136
Hua Wu, Taiyi Huang, Chengqing Zong, Bo Xu <i>Chinese Generation in a Spoken Dialogue Translation System</i>	1141
Alexander Yeh <i>Comparing Two Trainable Grammatical Relations Finders</i>	1146

Author Index

Aduriz, I.	1	Cerbah, F.	145
Agirre, E.	1	Cha, J.	1002
Aizawa, A.	397	Chan, G.	1121
Aizono, T.	404	Chan, S.	1121
Aldezabal, I.	1	Chao, G.	152
Alegria, I.	1	Charoenporn, T.	802
Allegrini, P.	8	Chen, C.	173
Amtrup, J.	982	Chen, H.	159, 166
Ananiadou, S.	530	Chen, K.	173
Ando, S.	926	Cheng, H.	306
Androutsopoulos, I.	15	Cheung, L.	1121
Aramaki, E.	906	Cho, H.	593
Armstrong, S.	1136	Church, K.	180
Arregi, X.	1	Ciaramita, M.	187
Arriola, J.	1	Clark, S.	194
Artola, X.	1	Collier, N.	201
Asahara, M.	21	Cowie, J.	1007
Atwell, E.	1021	Cristea, D.	208
Avgustinova, T.	28	Crocker, M.	111
Baldwin, T.	35	Daelemans, W.	327, 857
Bangalore, S.	42, 369	Daille, B.	215
Barg, P.	49	Dale, R.	15
Bateman, J.	474	Dang, H.	1011
Benedí, J.	55	van Deemter, K.	222
Bod, R.	62, 69	Déjean, H.	229, 857
Boguraev, B.	76	De Pauw, G.	236
Boguslavsky, I.	83, 987	Duan, L.	1031
Boitet, C.	768	Dyer, M.	152
Bond, F.	90	Dymetman, M.	243, 1016
Borin, L.	97	Ebert, C.	250
Branco, A.	104	Ecker, D.	1031
Brants, T.	111	Eisner, J.	257
Bratt, E.	670	Elliott, J.	1021
Brockett, C.	390, 822	Eriguchi, Y.	1106
Brown, R.	125	Fabre, C.	264
Brun, C.	132	Fakotakis, N.	808, 836
Bröker, N.	118	Flournoy, R.	1066
Buráňová, E.	139	Foth, K.	1026
Byron, D.	919	Francez, N.	1076
Cahill, L.	1126	Franz, A.	1031
Calder, J.	992	Frederking, R.	1041
Carberry, S.	740	Frid, N.	83, 987
Carl, M.	997	Furuse, O.	621
Ceplová, M.	278	Gawron, J.	670

Gillis, S.	327	Kanayama, H.	411
Goecke, K.	1091	Kang, I.	418
Gojenola, K.	1	Kaplan, R.	425
Goldwater, S.	670	Kaufmann, S.	1066
Görz, G.	1116	Kawahara, D.	432
Grigoriev, N.	987	Kawahara, T.	467
Grigorieva, S.	987	Kearns, M.	502
Grishman, R.	537, 940	Kiefer, B.	1046
Hahn, U.	271	Kilbury, J.	49
Hajičová, E.	139, 278	Kim, B.	1051
Hakkani-Tür, D.	285	Kim, G.	418
Hana, J.	474	Kim, J.	1096
Harabagiu, S.	292	Kim, S.	439
Hartley, T.	474	Kim, Y.	954
Hasida, K.	892	Kimura, K.	341
Hatzivassiloglou, V.	299	Kinyon, A.	446
Helfrich, A.	1036	Kipper, K.	1011
Henschel, R.	306	Klärner, M.	1116
Hirakawa, H.	313, 341	Ko, K.	1121
Hisamitsu, T.	320	Ko, Y.	453
Hockey, B.	670	Koeling, R.	857
Hogan, C.	1041	Kokkinakis, G.	808, 836
Hong, S.	383	Koller, A.	460
Horiguchi, K.	1031	Komatani, K.	467
Hoste, V.	327	König, E.	1056
Hovy, E.	495	Koontz, E.	1031
Huang, T.	1141	Kosaka, M.	537
Hurst, M.	334	Koyama, T.	878
Huttunen, S.	940	Kracht, M.	250
Ichimura, Y.	341	Kreidlin, L.	83, 987
Ide, N.	208	Krieger, H.	1046
Imamura, K.	572	Kruijff, G.	474
Inui, K.	348	Kruijff-Korbayová, I.	474
Inui, T.	348	Krymowski, Y.	857
Iomdin, L.	83	Kuessner, U.	843
Isahara, H.	509, 565, 754, 871, 885	Kurohashi, S.	432, 906
Itoh, N.	558	Lai, T.	1121
Jacquemin, C.	264	Lee, G.	1002, 1051
James, F.	670	Lee, J.	544
Johnson, M.	187, 355	Lee, S.	481
Johnston, M.	362, 369	Lepage, Y.	488
Jones, D.	376	Lezius, W.	1056
Jung, S.	383	Lim, H.	1096
Kacmarcik, G.	390, 822	Lin, C. Y.	495
Kageura, K.	397	Lin, C. J.	159
Kaji, H.	404	Litman, D.	502
Kaji, N.	432	Ludovik, Y.	1007, 1061

Lun, C.	1121	Ohara, H.	572
Lux, V.	243	Oka, M.	878
Ma, Q.	509, 565, 871	Okumura, M.	1071
Maiorano, S.	292	Ono, K.	313
Malavazos, C.	516	Osborne, M.	586
Marcu, D.	208, 523	Paek, E.	383
Maritxalar, A.	1	Paik, K.	90
Masuichi, H.	1066	Palmer, M.	1011
Matsumoto, Y.	21, 933	Park, J.	593
Matsuo, Y.	698	Paşca, M.	292
Maynard, D.	530	Pereira, C.	829
McKeown, K.	719	Perrier, G.	600
McNaught, J.	795	Peters, S.	1066
Megerdoomian, K.	982	Petitpierre, D.	1136
Menzel, W.	1026	Piperidis, S.	516
Merlo, P.	815	Pirrelli, V.	8
Meyers, A.	537	Piwek, P.	607
Mimura, K.	926	Pla, F.	614
Ming, J.	789	Planas, E.	621
Mitsubishi, Y.	411	Poesio, M.	306, 899
Miyauchi, T.	878	Pogodalla, S.	628
Molina, A.	614	Pop, H.	1026
Molina-Salgado, H.	1007	Potipiti, T.	802
Montemagni, S.	8	Poutsma, A.	635
Moon, K.	544	Power, R.	222, 642
Morawietz, F.	551	Prescher, D.	649
Mori, S.	558	Prieto, N.	614
Morimoto, Y.	404	Punyakankok, V.	857
Murata, M.	509, 565, 871, 885	Quint, J.	656
Music, B.	1036	Rambow, O.	42
Nakanishi, M.	579	Ramsay, A.	663, 775
Nakano, Y.	572	Ranta, A.	243
Nanba, H.	1071	Rayner, M.	670
Nasukawa, T.	334	Rehm, M.	1091
Neff, M.	76	Riezler, S.	649
Nelken, R.	1076	Rim, H.	481, 1096
Ney, H.	850, 1081, 1086, 1131	Roark, B.	355
Niehren, J.	460	Romacker, M.	271
Nießen, S.	1081	Rooth, M.	649
Nirenburg, S.	1007, 1111	Roth, D.	857
Nishimura, M.	558	Ruland, T.	677
Niwa, Y.	320	Rusk, G.	376
Nobata, C.	201	Ryu, W.	1096
Nobesawa, S.	579	Sagalova, I.	83
Och, F.	1086	Saito, H.	579
Oflazer, K.	285	Saito, Y.	341
Ogino, S.	558	Samuelsson, C.	684

Sánchez, J.	55	Torisawa, K.	411
Sarasola, K.	1	Tsai, J.	166
Sarkar, A.	691	Tsai, S.	166
Sasaki, Y.	698	Tseng, S.	864
Sassano, M.	705	Tsuji, K.	397
Schiehlen, M.	712, 1101	Tsujii, J.	201, 320, 411, 481
Schiffman, B.	719	Tür, G.	285
Schmid, H.	726	Uchida, K.	1031
Schneider, K.	733	Uchimoto, K.	509, 565, 754, 871
Schroeder, L.	740	Ueda, Y.	878
Schröder, I.	1026	Urkia, M.	1
Schulte im Walde, S.	726, 747	Uszkoreit, H.	28
Sekine, S.	754, 761, 871, 1106	Utiyama, M.	885, 892
Seo, J.	453	Utsuro, T.	705
Sérasset, G.	768	Vieira, R.	899
Seville, H.	663, 775	Vogel, S.	1131
Sgall, P.	139	Waibel, A.	968
Sharoff, S.	474	Walker, D.	1136
Sheremetyeva, S.	1111	Walker, M.	502
Shimohata, S.	782	Watanabe, H.	558, 906, 913
Sicilia-Garcia, E.	789	Watson, C.	829
Siegel, M.	1046	Wedekind, J.	425
Sierra, G.	795	Weir, D.	194
Sin, K.	1121	Whyte, B.	1021
Singh, S.	502	Wiebe, J.	299
Sizov, V.	83	Wolters, M.	919
Skoumalová, H.	474	Wu, H.	1141
Smith, F.	789	Xu, B.	1141
Sokolova, L.	474	Yamabana, K.	926
Song, M.	439, 954	Yamamoto, K.	933
Sornlertlamvanich, V.	802	Yamasaki, N.	404
Spilker, J.	1116	Yangarber, R.	940
Stamatatos, E.	808	Yeh, A.	947, 1146
Staykova, K.	474	Yoon, J.	439, 954
Stevenson, S.	815	Yoshimura, Y.	313
Suzuki, H.	390, 822	van Zaanen, M.	961
T'sou, B.	1121	Zacharski, R.	1061
Tablan, V.	208	Zajac, R.	982
Tanaka, H.	35	Zechner, K.	968
Tapanainen, P.	940	Zeman, D.	691
Teich, E.	474, 829	Zhou, Q.	975
Tendeau, F.	1016	Zong, C.	1141
Thanopoulos, A.	836		
Tiberius, C.	1126		
Tidhar, D.	843		
Tillmann, C.	850		
Tjong Kim Sang, E.	857		

