Lecture Notes in Artificial Intelligence 2166

Subseries of Lecture Notes in Computer Science
Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science
Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer
*Berlin*
*Heidelberg*
*New York*
*Barcelona*
*Hong Kong*
*London*
*Milan*
*Paris*
*Tokyo*

Václav Matoušek Pavel Mautner
Roman Mouček Karel Taušer (Eds.)

# Text, Speech and Dialogue

4th International Conference, TSD 2001
Železná Ruda, Czech Republic, September 11-13, 2001
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Václav Matoušek
Pavel Mautner
Roman Mouček
Karel Taušer
University of West Bohemia in Plzeň, Faculty of Applied Sciences
Dept. of Computer Science and Engineering
Univerzitní 22, 306-14 Plzeň, Czech Republic
E-mail: {matousek/mautner/moucek/tauser}@kiv.zcu.cz

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Text, speech and dialogue : 4th international conference ; proceedings /
TSD 2001, Železná Ruda, Czech Republic, September 11 - 13, 2001.
Václav Matoušek ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ;
Hong Kong ; London ; Milan ; Paris ; Tokyo : Springer, 2001
(Lecture notes in computer science ; Vol. 2166 : Lecture notes in
artificial intelligence)
ISBN 3-540-42557-8

CR Subject Classification (1998): I.2.7., H.3, H.4, I.7

ISBN 3-540-42557-8 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

© Springer-Verlag Berlin Heidelberg 2001
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Steingräber Satztechnik GmbH, Heidelberg
Printed on acid-free paper SPIN: 10840355 06/3142 5 4 3 2 1 0

# Preface

The International Conference TSD 2001, the fourth event in the series on Text, Speech, and Dialogue, which originated in 1998, presents state–of–the–art technology and recent achievements in the field of natural language processing. But, this year's TSD 2001 is organized as the "real" conference involving all components belonging to such an event – invited talks given by top-class researchers and plenary and problem oriented sessions are preceeded by four tutorials evolving around all problem areas of human–machine interaction.

The conference declares its intent to be an interdisciplinary forum, which intertwines research in speech and language processing as well as research in the Eastern and Western hemispheres. We feel that the mixture of different approaches and applications gives a great opportunity to get acquainted with the current activities in all aspects of language communication and to witness the amazing vitality of researchers from the former East Block countries. The financial support of ISCA (International Speech Communication Association) enables the wide attendance of researchers from all active regions of the world.

This book contains a collection of all the papers presented at the international conference organized by the Faculty of Applied Sciences of the University of West Bohemia in Pilsen in collaboration with the Faculty of Informatics, Masaryk University in Brno, and held in the beautiful setting of Železná Ruda (Czech Republic), September 11–13, 2001. A total of 59 accepted papers out of 117 submitted, contributed by 122 authors (47 from Central Europe, 11 from Eastern Europe, 54 from Western Europe, 5 from America, and 5 from Asia) are included in these conference proceedings.

We would like to gratefully thank the invited speakers and the authors of the papers for their valuable contributions, the ISCA for its financial support, and Prof. Ježek, Dean of the Faculty of Applied Sciences, for greeting the conference on behalf of the University of West Bohemia.

Last but not least, we would like to express our gratitude to the authors for their strive to provide the papers in time, to members of the Program Committee for their careful reviews, to the editors for their hard work in preparing these proceedings, and to the members of the Local Organizing Committee for their enthusiasm in organizing the conference.

June 2001 H. Hermansky and V. Matoušek

# Organization

TSD 2001 was organized by the Faculty of Applied Sciences, University of West Bohemia in Plzeň (Pilsen), in cooperation with the Faculty of Informatics, Masaryk University in Brno, Czech Republic. The conference Web–page is located at URL:
`http://www-kiv.zcu.cz/events/tsd2001/.`

## Program Committee

Frederick Jelinek (USA), *general chair*
Hynek Hermansky (USA), *executive chair*
Eneko Agirre (Spain)
Geneviève Baudoin (France)
František Čermák (Czech Republic)
Attila Ferencz (Romania)
Eva Hajičová (Czech Republic)
Patrick Hanks (UK)
Adam Kilgarriff (UK)
Ivan Kopeček (Czech Republic)
Václav Matoušek (Czech Republic)
Rosmaund Moon (UK)
Elmar Nöth (Germany)
Karel Pala (Czech Republic)
Nikola Pavešić, (Slovenia)
Vladimír Petkevič (Czech Republic)
Josef Psutka (Czech Republic)
Léon J. M. Rothkrantz (The Netherlands)
Ernst Günter Schukat-Talamazzini (Germany)
Pavel Skrelin (Russia)
Taras Vintsiuk (Ukraine)
Yorick Wilks (UK)

## Local Organizing Committee

Václav Matoušek *(chair)*, Helena Benešová *(secretary)*, Kamil Ekštein, Jana Hesová, Svatava Kindlová, Jana Klečková, Ivan Kopeček, Jana Krutišová, Josef Mašek, Pavel Mautner, Roman Mouček, Ondřej Rohlík, Pavel Slavík, Petr Sojka, Karel Taušer

**Supported by:**

International Speech Communication Association (ISCA)

# Table of Contents

**Text**


Knowledge Sources for Word Sense Disambiguation . . . . . 1
*E. Agirre and D. Martinez*

The Current Status of the Prague Dependency Treebank . . . . . 11
*E. Hajičová, J. Hajič, B. Hladká, M. Holub, P. Pajas, V. Řezníčková, and P. Sgall*

Language Corpora: The Czech Case . . . . . 21
*F. Čermák*

A Hybrid Logic Formalization of Information Structure Sensitive Discourse Interpretation . . . . . 31
*G.-J.M. Kruijff and I. Kruijff-Korbayová*

The Possibilities of Automatic Detection/Correction of Errors in Tagged Corpora: A Pilot Study on a German Corpus . . . . . 39
*K. Oliva*

Grammatical Agreement and Automatic Morphological Disambiguation of Inflectional Languages . . . . . 47
*V. Petkevič*

Modelling Semantic Association and Conceptual Inheritance for Semantic Analysis . . . . . 54
*P. Vaillant*

Influence of Conditional Independence Assumption on Verb Subcategorization Detection . . . . . 62
*K. Kermanidis, M. Maragoudakis, N. Fakotakis, and G. Kokkinakis*

Morphological Guesser of Czech Words . . . . . 70
*J. Hlaváčová*

Clustering Technique Based on Semantics for Definite Description Resolution . . . 76
*R. Muñoz and M. Palomar*

Pronoun Resolution in Spanish from Full Parsing . . . . . 84
*M. Saiz-Noeda, M. Palomar, and L. Moreno*

A Method of Accurate Robust Parsing of Czech . . . . . 92
*V. Kuboň and M. Plátek*

A New Czech Morphological Analyser ajka . . . . . 100
*R. Sedláček and P. Smrž*

Finding Semantically Related Words in Large Corpora . . . . . 108
*P. Smrž and P. Rychlý*

Syntactic-Based Methods for Measuring Word Similarity . . . . . 116
*P. Gamallo, C. Gasperin, A. Agustini, and G.P. Lopes*

The ATRACT Workbench:
Automatic Term Recognition and Clustering for Terms . . . . . 126
*H. Mima, S. Ananiadou, and G. Nenadić*

Augmented Auditory Representation of e-Texts for Text-to-Speech Systems . . . . . 134
*G. Xydas and G. Kouroupetroglou*

Enhancing the Valency Dictionary of Czech Verbs:
Tectogrammatical Annotation . . . . . 142
*H. Skoumalová, M. Straňáková-Lopatková, and Z. Žabokrtský*

An Interactive Graph Based Legal Information Retrieval System . . . . . 150
*P. Quaresma and I.P. Rodrigues*

Text Segmentation into Paragraphs Based on Local Text Cohesion . . . . . 158
*I.A. Bolshakov and A. Gelbukh*

Building a Digital Collection of Web-Pages:
Access and Filtering Information with Textual Expansion . . . . . 167
*O. Larouk*

Three Approaches to Word Sense Disambiguation for Czech . . . . . 174
*R. Král*

Method for WordNet Enrichment Using WSD . . . . . 180
*A. Montoyo, M. Palomar, and G. Rigau*

**Speech**

Human Speech Perception: Some Lessons from Automatic Speech Recognition . . 187
*H. Heřmanský*

Speech Technology in Reality – Applications, Their Challenges and Solutions . . . 197
*H. Mangold*

Generative Phoneme-Threephone Model for ASR . . . . . 201
*T.K. Vintsiuk*

Algebraic Models of Speech Segment Databases . . . . . 208
*I. Kopeček*

Bayesian Noise Compensation of Time Trajectories of Spectral Coefficients for Robust Speech Recognition . . . . . 214
*I. Potamitis, N. Fakotakis, and G. Kokkinakis*

The Influence of a Filter Shape in Telephone-Based Recognition Module Using PLP Parameterization . . . . . 222
*J. Psutka, L. Müller, and J.V. Psutka*

System for Speech Communication through Internet . . . . . 229
*A. Machovikov and K. Stolyarov*

Phoneme Based ASR System for Slovak SpeechDat Database . . . . . 237
*M. Marcinek, J. Juhar, and A. Cizmar*

Large Vocabulary Continuous Speech Recognizer for Slovenian Language . . . . . 242
*A. Žgank, Z. Kačič, and B. Horvat*

A Nonlinearized Discriminant Analysis and Its Application to Speech Impediment Therapy . . . . . 249
*A. Kocsor, L. Tóth, and D. Paczolay*

Speech Recognition Issues for Dutch Spoken Document Retrieval . . . . . 258
*R. Ordelman, A. van Hessen, and F. de Jong*

Multichannel Sound Acquisition with Stress Situations Determination for Medical Supervision in a Smart House . . . . . 266
*D. Istrate and E. Castelli*

Two-Pass Recognition of Czech Speech Using Adaptive Vocabulary . . . . . 273
*P. Ircing and J. Psutka*

Simulation of Speaking Styles with Adapted Prosody . . . . . 278
*H. Kruschke*

The Utterance Type and the Size of the Intonation Unit . . . . . 285
*N. Volskaya and P. Skrelin*

Estimation of Boundaries between Speech Units Using Bayesian Changepoint Detectors . . . . . 291
*R. Čmejla and P. Sovka*

Data Driven Design of Filter Bank for Speech Recognition . . . . . 299
*L. Burget and H. Heřmanský*

Minimization of Transition Noise and HNM Synthesis in Very Low Bit Rate Speech Coding . . . . . 305
*P. Motlíček, G. Baudoin, J. Černocký, and G. Chollet*

Chinese Radicals and Top Ontology in EuroWordNet . . . . . 313
*S.H.S. Wong and K. Pala*

Towards a Dynamic Adjustment of the Language Weight . . . . . 323
*G. Stemmer, V. Zeissler, E. Nöth, and H. Niemann*

Speech Recognition Based on Feature Extraction
with Variable Rate Frequency Sampling . . . . . 329
*I. Potamitis, N. Fakotakis, and G. Kokkinakis*

The Phonectic SMS Reader . . . . . 334
*J. Gros, F. Mihelič, N. Pavešić, M. Žganec, A. Mihelič, M. Knez, A. Merčun, and D. Škerl*

**Dialogue**

Research Issues for the Next Generation Spoken Dialogue Systems Revisited . . . . 341
*E. Nöth, M. Boros, J. Fischer, F. Gallwitz, J. Haas, R. Huber, H. Niemann, G. Stemmer, and V. Warnke*

Determining User Interface Semantics Using Communicating Agents . . . . . 349
*L. Ton, L.J.M. Rothkrantz*

Agent-Based Adaptive Interaction and Dialogue Management Architecture
for Speech Applications . . . . . 357
*M. Turunen and J. Hakulinen*

Dialogue for Web Search Utilizing Automatically Acquired
Domain Knowledge . . . . . 365
*U. Kruschwitz*

A Development Tool for Spoken Dialogue Systems and Its Evaluation . . . . . 373
*S. Kogure and S. Nakagawa*

Using XML for Representing Domain Dependent Knowledge in Dialogos . . . . . 381
*G. Boella, R. Damiano, M. Danieli, and L. Lesmo*

Dialogue Generation of Program Source Codes . . . . . 388
*P. Nygrýn and R. Batůšek*

Semi-automatic Identification of Danish Discourse Deictics . . . . . 396
*C. Navarretta*

Language Understanding Using Two-Level Stochastic Models
with POS and Semantic Units . . . . . 403
*F. Pla, A. Molina, E. Sanchis, E. Segarra, F. García*

Shallow Processing and Cautious Incrementality in a Dialogue System Front End:
Two Steps towards Robustness and Reactivity . . . . . 410
*T. Lager*

Creation of a Corpus of Training Sentences
Based on Automated Dialogue Analysis . . . . . 418
*J. Schwarz and V. Matoušek*

Dialogue Manager in Dialogue System for Visually Impaired Programmers . . . . . 427
*K. Taušer, V. Matoušek, and R. Mouček*

Help System in Dialogue Grammar Based Programming Environment
for Visually Impaired Programmers . . . . . 435
*R. Mouček, K. Taušer, and P. Nygrýn*

Automatic Generation of Dialogue Interfaces for Web-Based Applications . . . . . 443
*L. Bártek*

**Author Index** . . . . . 451