

Lecture Notes in Artificial Intelligence

1902

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

Singapore

Tokyo

Petr Sojka Ivan Kopeček
Karel Pala (Eds.)

Text, Speech and Dialogue

Third International Workshop, TSD 2000
Brno, Czech Republic, September 13-16, 2000
Proceedings



Springer

Series Editors

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

Volume Editors

Petr Sojka

Masaryk University, Faculty of Informatics
Department of Programming Systems and Communication
Botanická 68a, 602 00 Brno, Czech Republic
E-mail: sojka@informatics.muni.cz

Ivan Kopeček

Karel Pala

Masaryk University, Faculty of Informatics, Department of Information Technologies
Botanická 68a, 602 00 Brno, Czech Republic
E-mail: {kopecek/pala}@informatics.muni.cz

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Text, speech and dialogue : third international workshop ; proceedings /
TSD 2000, Brno, Czech Republic, September 13 - 16, 2000. Petr Sojka
... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ;
London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2000
(Lecture notes in computer science ; 1902 : Lecture notes in
artificial intelligence)
ISBN 3-540-41042-2

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

ISBN 3-540-41042-2 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
© Springer-Verlag Berlin Heidelberg 2000
Printed in Germany

Typesetting: Camera-ready by editors from source files provided by authors, data conversion by Steingraber
Satztechnik GmbH, Heidelberg
Printed on acid-free paper SPIN: 10722638 06/3142 5 4 3 2 1 0

Preface

The workshop series on Text, Speech and Dialogue originated in 1998 with the first TSD 1998 held in Brno, Czech Republic. This year's TSD 2000, already the third in the series, returns to Brno and to its organizers from the Faculty of Informatics at the Masaryk University. As shown by the ever growing interest in TSD series, this annual workshop developed into the prime meeting of speech and language researchers from both sides of the former Iron Curtain, which provides a unique 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. Thanks need to be extended to all who continue to make the TSD workshop series such a success: first, to the authors themselves, without whom TSD 2000 would not exist; next, to all organizations that support TSD 2000, among them the International Speech Communication Association, the Faculty of Informatics at the Masaryk University in Brno and the Faculty of Applied Sciences, West Bohemia University in Plzeň; and last but not least, to the organizers and members of the Program Committee who spent much effort to make TSD 2000 success and who reviewed 131 contributions submitted from all corners of the world and accepted 75 out of them for presentation at the workshop. This book is evidence of the success of all involved.

June 2000

Hynek Hermansky

Organization

TSD 2000 was organized by the Faculty of Informatics, Masaryk University, in cooperation with the Faculty of Applied Sciences, University of West Bohemia in Plzeň. The workshop's Web page is located at <http://www.fi.muni.cz/tsd2000/>.

Program Committee

Jelinek, Frederick (USA), *general chair*
Hermansky, Hynek (USA), *executive chair*
Baudoin, Geneviève (France)
Čermák, František (Czech Republic)
Ferencz, Attila (Romania)
Hajičová, Eva (Czech Republic)
Hanks, Patrick (Great Britain)
Kilgarriff, Adam (Great Britain)
Kopeček, Ivan (Czech Republic)
Kučera, Karel (Czech Republic)
Matoušek, Václav (Czech Republic)
Moon, Rosamund (Great Britain)
Nöth, Elmar (Germany)
Pala, Karel (Czech Republic)
Pavesić, Nikola (Slovenia)
Petkevič, Vladimír (Czech Republic)
Psutka, Josef (Czech Republic)
Schukat-Talamazzini, E. Günter (Germany)
Skrelin, Pavel (Russia)
Vintsiuk, Taras (Ukraine)
Wilks, Yorick (Great Britain)

Organizing Committee

Luděk Bártek, Robert Batůšek, Pavel Gaura, Aleš Horák, Dagmar Janoušková, Dana Komárková (*secretary*), Ivan Kopeček (*co-chair*), Václav Matoušek, Pavel Nygrýn, Karel Pala (*co-chair*), Pavel Smrž, Petr Sojka (*Proceedings*), Jan Staudek, Marek Veber, Eva Žáčková

Supported by:

International Speech Communication Association

Table of Contents

I Text

The Linguistic Basis of a Rule-Based Tagger of Czech	3
<i>Karel Oliva (University of Saarland), Milena Hnátková, Vladimír Petkevič, Pavel Květoň (Charles University, Prague)</i>	
Harnessing the Lexicographer in the Quest for Accurate Word Sense Disambiguation	9
<i>David Tugwell, Adam Kilgarriß (ITRI, University of Brighton)</i>	
An Integrated Statistical Model for Tagging and Chunking Unrestricted Text	15
<i>Ferran Pla, Antonio Molina, Natividad Prieto (University of Valencia)</i>	
Extending Bidirectional Chart Parsing with an Stochastic Model	21
<i>Alicia Ageno, Horacio Rodríguez (Universidad Politècnica de Catalunya)</i>	
Ensemble of Classifiers for Noise Detection in PoS Tagged Corpora	27
<i>Harald Berthelsen (Telia Promotor), Beáta Megyesi (Stockholm University)</i>	
Towards a Dynamic Syntax for Language Modelling	33
<i>David Tugwell (University of Brighton)</i>	
A Word Analysis System for German Hyphenation, Full Text Search, and Spell Checking, with Regard to the Latest Reform of German Orthography	39
<i>Gabriele Kodydek (Vienna University of Technology)</i>	
Automatic Functor Assignment in the Prague Dependency Treebank	45
<i>Zdeněk Žabokrtský (Czech Technical University, Prague)</i>	
Categories, Constructions, and Dependency Relations	51
<i>Geert-Jan M. Kruijff (Charles University, Prague)</i>	
Local Grammars and Parsing Coordination of Nouns in Serbo-Croatian ...	57
<i>Goran Nenadić (University of Belgrade)</i>	
Realization of Syntactic Parser for Inflectional Language Using XML and Regular Expressions	63
<i>Marek Trabalka and Mária Bieliková (Slovak University of Technology, Bratislava)</i>	
A Rigoristic and Automated Analysis of Texts Applied to a Scientific Abstract by Mark Sergot and Others	69
<i>Gregers Koch (Copenhagen University)</i>	

Evaluation of Tectogrammatical Annotation of PDT.....	75
<i>Eva Hajičová, Petr Pajas (Charles University, Prague)</i>	
Probabilistic Head-Driven Chart Parsing of Czech Sentences.....	81
<i>Pavel Smrž, Aleš Horák (Masaryk University in Brno)</i>	
Aggregation and Contextual Reference in Automatically Generated Instructions	87
<i>Geert-Jan M. Kruijff, Ivana Kruijff-Korbayová (Charles University, Prague)</i>	
Information Retrieval by Means of Word Sense Disambiguation	93
<i>Luis Alfonso Ureña (University of Jaén), José María Gómez Hidalgo, Manuel de Buenaga (Universidad Europea- CEES, Madrid)</i>	
Statistical Parameterisation of Text Corpora	99
<i>Gregory Y. Martynenko, Tatiana Y. Sherstinova (St. Petersburg State University)</i>	
An Efficient Algorithm for Japanese Sentence Compaction Based on Phrase Importance and Inter-Phrase Dependency.....	103
<i>Rei Oguro, Kazuhiko Ozeki, Kazuyuki Takagi (The University of Electro- Communications, Tokyo), Yujie Zhang (ATR Spoken Language Trans- lation Research Laboratories, Kyoto)</i>	
Word Senses and Semantic Representations	109
<i>Karel Pala (Masaryk University in Brno)</i>	
Automatic Tagging of Compound Verb Groups in Czech Corpora	115
<i>Eva Žáčková, Luboš Popelínský, Miloslav Nepil (Masaryk University in Brno)</i>	
Sensitive Words and Their Application to Chinese Processing.....	121
<i>Fuji Ren (Hiroshima City University), Jian-Yun Nie (University of Montreal)</i>	
Testing a Word Analysis System for Reliable and Sense-Conveying Hyphenation and Other Applications	127
<i>Martin Schönhacker, Gabriele Kodydek (Vienna University of Technol- ogy)</i>	
The Challenge of Parallel Text Processing	133
<i>Milena Slavcheva (Bulgarian Academy of Science)</i>	
Selected Types of Pg-Ambiguity: Processing Based on Analysis by Reduction	139
<i>Markéta Straňáková (Charles University, Prague)</i>	

Cohesive Generation of Syntactically Simplified Newspaper Text	145
<i>Yvonne Canning, John Tait, Jackie Archibald, Ros Crawley (University of Sunderland)</i>	
TEA: A Text Analysis Tool for the Intelligent Text Document Filtering . . .	151
<i>Jan Žižka, Aleš Bourek, Luděk Frey (Masaryk University in Brno)</i>	
Competing Patterns for Language Engineering	157
<i>Petr Sojka (Masaryk University in Brno)</i>	

II Speech

Recognition and Labelling of Prosodic Events in Slovenian Speech	165
<i>France Mihelič, Jerneja Gros (University of Ljubljana), Elmar Nöth, Volker Warnke (University of Erlangen-Nürnberg)</i>	
Rules for Automatic Grapheme-to-Allophone Transcription in Slovene	171
<i>Jerneja Gros, France Mihelič, Simon Dobrišek, Tomaž Erjavec, Mario Žganec (University of Ljubljana)</i>	
An Adaptive and Fast Speech Detection Algorithm	177
<i>Dragos Burileanu, Lucian Pascalin, Corneliu Burileanu (University of Bucharest), Mihai Puchiu (Graphco Technologies, Romania)</i>	
Optimal Pitch Path Tracking for More Reliable Pitch Detection	183
<i>Petr Motlíček, Jan Černocký (VUT Brno)</i>	
FlexVoice: A Parametric Approach to High-Quality Speech Synthesis	189
<i>György Balogh, Ervin Dobler, Tamás Gröbner, Béla Smolics, Csaba Szepesvári (Mindmaker Ltd., Budapest)</i>	
The Continuous and Discontinuous Styles in Czech TTS	195
<i>Tomáš Duběda, Jiří Hanika (Charles University, Prague)</i>	
Automatic Speech Segmentation	
with the Application of the Czech TTS System	201
<i>Petr Horák, Betty Hesounová (Academy of Sciences of the Czech Republic, Prague)</i>	
Speaker Identification Using Autoregressive Hidden Markov Models	
and Adaptive Vector Quantisation	207
<i>Eugeniy E. Bovbel, Igor E. Kheidorov, Michael E. Kotlyar (Belarussian State University)</i>	
Morpheme Based Language Models for Speech Recognition of Czech	211
<i>William Byrne (Johns Hopkins University), Jan Hajič, Pavel Krbeč (Charles University, Prague), Pavel Ircing, Josef Psutka (University of West Bohemia in Pilsen)</i>	

A Large Czech Vocabulary Recognition System for Real-Time Applications	217
<i>Jan Nouza (Technical University of Liberec)</i>	
Building a New Czech Text-to-Speech System Using Triphone-Based Speech Units	223
<i>Jindřich Matoušek (University of West Bohemia in Pilsen)</i>	
Acoustic and Perceptual Properties of Syllables in Continuous Speech as a Function of Speaking Rate	229
<i>Hisao Kuwabara, Michinori Nakamura (Teikyo University of Science and Technology)</i>	
NL-Processor and Linguistic Knowledge Base in a Speech Recognition System	237
<i>Michael G. Malkovsky, Alexey V. Subbotin (Moscow State University)</i>	
Russian Phonetic Variability and Connected Speech Transcription	243
<i>Vladimir I. Kuznetsov, Tatiana Y. Sherstinova (St. Petersburg State University)</i>	
Database Processing for Spanish Text-to-Speech Synthesis	248
<i>Juan Francisco Gómez-Mena, M. Cardo, José Luis Madrid, C. Prades (Polytechnic University, Madrid)</i>	
Topic-Sensitive Language Modelling	253
<i>Mirjam Sepesy Maučec, Zdravko Kačič (University of Maribor)</i>	
Design of Speech Recognition Engine	259
<i>Luděk Müller, Josef Psutka, Luboš Šmídl (University of West Bohemia in Pilsen)</i>	
Combining Multi-band and Frequency-Filtering Techniques for Speech Recognition in Noisy Environments	265
<i>Peter Jančovič, Ji Ming, Philip Hanna, Darryl Stewart, Jack Smith (The Queen's University of Belfast)</i>	
Allophone- and Suballophone-Based Speech Synthesis System for Russian .	271
<i>Pavel Skrelin (St. Petersburg State University)</i>	
Diphone-Based Unit Selection for Catalan Text-to-Speech Synthesis	277
<i>Roger Gvaus i Térmens, Ignasi Iriondo Sanz (University Ramon Llull, Barcelona)</i>	
Analysis of Information in Speech and Its Application in Speech Recognition	283
<i>Sachin S. Kajarekar (Oregon Graduate Institute of Science and Tech- nology), Hynek Hermansky (International Computer Science Institute, Berkeley)</i>	

What Textual Relationships Demand Phonetic Focus?	289
<i>Sofia Gustafson-Čapková, Jennifer Spenader (Stockholm University)</i>	
Speaker Identification Using Kalman Cepstral Coefficients	295
<i>Zdeněk Švenda, Vlasta Radová (University of West Bohemia in Pilsen)</i>	
Belarussian Speech Recognition Using Genetic Algorithms	301
<i>Eugene I. Bovbel, Dzmitry V. Tsishkou (Belarussian State University)</i>	
A Discriminative Segmental Speech Model and Its Application to Hungarian Number Recognition	307
<i>László Tóth, András Kocsor, Kornél Kovács (University of Szeged)</i>	
Comparison of Frequency Bands in Closed Set Speaker Identification Performance	314
<i>Özgür Devrim Orman (TÜBİTAK, Kocaeli), Levent Arslan (Boğaziçi University, İstanbul)</i>	
Recording and Annotation of the Czech Speech Corpus	319
<i>Vlasta Radová, Josef Psutka (University of West Bohemia in Pilsen)</i>	

III Dialogue

A Text Based Talking Face	327
<i>Léon J.M. Rothkrantz, Ania Wojdel (Delft University of Technology)</i>	
Dialogue Control in the Alparon System	333
<i>Léon J.M. Rothkrantz, Robert J. van Vark, Alexandra Peters, Niels A. Andeweg (Delft University of Technology)</i>	
ISIS: Interaction through Speech with Information Systems	339
<i>Afzal Ballim, Jean-Cédric Chappelier, Martin Rajman, Vincenzo Pal- lotta (EPFL, Lausanne)</i>	
Centering-Based Anaphora Resolution in Danish Dialogues	345
<i>Costanza Navarretta (Center for Language Technology, Copenhagen)</i>	
Some Improvements on the IRST Mixed Initiative Dialogue Technology . . .	351
<i>Cristina Barbero, Daniele Falavigna, Roberto Gretter, Marco Orlandi, Emanuele Pianta, (ITC-irst, Trento)</i>	
Dictionary-Based Method for Coherence Maintenance in Man-Machine Dialogue with Indirect Antecedents and Ellipses	357
<i>Alexander Gelbukh, Grigori Sidorov, Igor A. Bolshakov (Natural Lan- guage Laboratory, Mexico)</i>	

Reconstructing Conversational Games in an Obligation-Driven Dialogue Model	363
<i>Jörn Kreutel (SAIL LABS S.L., Spain)</i>	
Prosody Prediction from Tree-Like Structure Similarities	369
<i>Laurent Blin (IRISA-ENSSAT), Mike Edgington (SRI International)</i>	
A Speaker Authentication Module in TelCorreo	375
<i>Leandro Rodríguez Liñares (University of Ourense), Carmen García-Mateo (University of Vigo)</i>	
TelCorreo: A Bilingual E-mail Client over the Telephone	381
<i>Leandro Rodríguez Liñares (University of Ourense), Antonio Cardenal López, Carmen García Mateo, David Pérez-Piñar López, Eduardo Rodríguez Banga, Xabier Fernández Salgado (University of Vigo)</i>	
A Syntactical Model of Prosody as an Aid to Spoken Dialogue Systems in Italian Language	387
<i>Enzo Mumolo (University of Trieste)</i>	
What Do You Mean by “What Do You Mean”?	393
<i>Norihiro Ogata (Osaka University)</i>	
Simplified Processing of Elliptic and Anaphoric Utterances in a Train Timetable Information Retrieval Dialogue System	399
<i>Václav Matoušek (University of West Bohemia in Pilsen)</i>	
Pragmatic and Grammatical Aspects of the Development of Dialogue Strategies	405
<i>James Monaghan (University of Hertfordshire)</i>	
An Annotation Scheme for Dialogues Applied to Anaphora Resolution Algorithms	410
<i>Patricio Martínez-Barco, Manuel Palomar (University of Alicante)</i>	
Cooperative Information Retrieval Dialogues through Clustering	415
<i>Paulo Quaresma, Irene Pimenta Rodrigues (University Évora)</i>	
Acoustic Cues for Classifying Communicative Intentions in Dialogue Systems	421
<i>Michelina Savino, Mario Refice (Politecnico di Bari)</i>	
Active and Passive Strategies in Dialogue Program Generation	427
<i>Ivan Kopeček (Masaryk University in Brno)</i>	
Architecture of Multi-modal Dialogue System	433
<i>Martin Fuchs, Petr Hejda, Pavel Slavík (Czech Technical University, Prague)</i>	

The Utility of Semantic-Pragmatic Information and Dialogue-State for Speech Recognition in Spoken Dialogue Systems	439
<i>Georg Stemmer, Elmar Nöth, Heinrich Niemann (University of Erlangen-Nürnberg)</i>	
Word Concept Model for Intelligent Dialogue Agents	445
<i>Yang Li, Tong Zhang, Stephen E. Levinson (Beckman Institute, Urbana)</i>	
Author Index	451
Subject Index	453