

Petr Sojka Aleš Horák
Ivan Kopeček Karel Pala (Eds.)

Text, Speech and Dialogue

11th International Conference, TSD 2008
Brno, Czech Republic, September 8-12, 2008
Proceedings

Table of Contents

I Invited Papers

The Future of Text-Meaning in Computational Linguistics	3
<i>Graeme Hirst (University of Toronto, Canada)</i>	
Deep Lexical Semantics	13
<i>Jerry Hobbs (University of Southern California, United States)</i>	
Toward the Ultimate ASR Language Model	15
<i>Frederick Jelinek and Carolina Parada (John Hopkins University, Baltimore, United States)</i>	
Practical Prosody: Modeling Language Beyond the Words	17
<i>Elizabeth Shriberg (SRI International, United States)</i>	

II Text

Sentiment Detection Using Lexically-Based Classifiers	21
<i>Ben Allison (University of Sheffield, United Kingdom)</i>	
First Steps in Building a Verb Valency Lexicon for Romanian	29
<i>Ana-Maria Barbu (Romanian Academy, Bucharest, Romania)</i>	
Hyponymy Patterns: Semi-automatic Extraction, Evaluation and Inter-lingual Comparison	37
<i>Virginica Barbu Mititelu (Romanian Academy, Bucharest, Romania)</i>	
Temporal Issues and Recognition Errors on the Capitalization of Speech Transcriptions	45
<i>Fernando Batista (L²F & ISCTE, Lisboa, Portugal), Nuno Mamede (L²F & IST, Lisboa, Portugal), and Isabel Trancoso (L²F & ISCTE, Lisboa, Portugal)</i>	
Web-Based Lemmatisation of Named Entities	53
<i>Richárd Farkas (MTA-SZTE, Szeged, Hungary), Veronika Vincze, István Nagy (University of Szeged, Hungary), Róbert Ormándi, György Szarvas (MTA-SZTE, Szeged, Hungary), and Attila Almási (University of Szeged, Hungary)</i>	
Combining Multiple Resources to Build Reliable Wordnets	61
<i>Darja Fišer (University of Ljubljana, Slovenia) and Benoît Sagot (INRIA, Paris, France)</i>	

Active Tags for Semantic Analysis	69
<i>Ivan Habernal and Miloslav Konopík (University of West Bohemia, Pilsen, Czech Republic)</i>	
Spoken Requests for Tourist Information: Speech Acts Annotation	77
<i>Laura Hasler (University of Wolverhampton, United Kingdom)</i>	
Affix: Tool for Prefix Recognition	85
<i>Jaroslava Hlaváčová and Michal Hrušecký (Charles University in Prague, Czech Republic)</i>	
Variants and Homographs: Eternal Problem of Dictionary Makers	93
<i>Jaroslava Hlaváčová and Markéta Lopatková (Charles University in Prague, Czech Republic)</i>	
An Empirical Bayesian Method for Detecting Out of Context Words	101
<i>Sanaz Jabbari, Ben Allison, and Louise Guthrie (University of Sheffield, United Kingdom)</i>	
Semantic Classes in Czech Valency Lexicon: Verbs of Communication and Verbs of Exchange	109
<i>Václava Kettnerová, Markéta Lopatková, and Klára Hrstková (Charles University in Prague, Czech Republic)</i>	
A Comparison of Language Models for Dialog Act Segmentation of Meeting Transcripts	117
<i>Jáchym Kolář (University of West Bohemia, Pilsen, Czech Republic)</i>	
New Methods for Pruning and Ordering of Syntax Parsing Trees: Comparison and Combination	125
<i>Vojtěch Kovář, Aleš Horák, and Vladimír Kadlec (Masaryk University, Brno, Czech Republic)</i>	
Web People Search with Domain Ranking	133
<i>Zornitsa Kozareva (University of Alicante, Spain), Rumen Moraliyski (University of Beira Interior, Portugal), and Gaël Dias (University of Alicante, Spain)</i>	
Disyllabic Chinese Word Extraction Based on Character Thesaurus and Semantic Constraints in Word-Formation	141
<i>Sun Maosong, Xu Dongliang (Tsinghua University, Beijing, China), Benjamin K.Y. T'sou (City University of Hong Kong, China), and Lu Huaming (Beijing Information Science and Technology University, Beijing, China)</i>	
Statistical Word Sense Disambiguation in Contexts for Russian Nouns Denoting Physical Objects	153
<i>Olga Mitrofanova (St. Petersburg State University, Russia), Olga Lashevskaya (Institute of the Russian Language, Moscow, Russia), and Polina Panicheva (St. Petersburg State University, Russia)</i>	

Where Do Parsing Errors Come from: The Case of Spoken Estonian	161
<i>Kaili Müürisep and Helen Nigol (University of Tartu, Estonia)</i>	
Dealing with Small, Noisy and Imbalanced Data: Machine Learning or Manual Grammars?	169
<i>Adam Przepiórkowski (Polish Academy of Sciences & Warsaw University, Warsaw, Poland), Michał Marcińczuk (Wrocław University of Technology, Poland), and Łukasz Degórski (Polish Academy of Sciences, Warsaw, Poland)</i>	
Statistical Properties of Overlapping Ambiguities in Chinese Word Segmentation and a Strategy for Their Disambiguation	177
<i>Wei Qiao, Maosong Sun (Tsinghua University, Beijing, China), and Wolfgang Menzel (Hamburg University, Germany)</i>	
The Verb Argument Browser	187
<i>Bálint Sass (Pázmány Péter Catholic University, Budapest)</i>	
Thinking in Objects: Towards an Infrastructure for Semantic Representation of Verb-Centred Structures	193
<i>Milena Slavcheva (Bulgarian Academy of Sciences, Sofia, Bulgaria)</i>	
Improving Unsupervised WSD with a Dynamic Thesaurus	201
<i>Javier Tejada-Cárcamo (National Polytechnic Institute, Mexico City, Mexico & Peruvian Computer Society, Arequipa, Peru), Hiram Calvo, and Alexander Gelbukh (National Polytechnic Institute, Mexico City, Mexico)</i>	
Normalization of Temporal Information in Estonian	211
<i>Margus Treumuth (University of Tartu, Estonia)</i>	
Word Sense Disambiguation with Semantic Networks	219
<i>George Tsatsaronis, Iraklis Varlamis, and Michalis Vazirgiannis (Athens University of Economics and Business, Athens, Greece)</i>	
A Pattern-Based Framework for Uncertainty Representation in Ontologies	227
<i>Miroslav Vacura, Vojtěch Svátek (University of Economics, Prague, Czech Republic), and Pavel Smrž (Brno University of Technology, Czech Republic)</i>	
MSD Recombination for Statistical Machine Translation into Highly-Inflected Languages	235
<i>Jerneja Žganec-Gros and Stanislav Gruden (Alpineon d.o.o., Ljubljana, Slovenia)</i>	
A Computational Framework to Integrate Different Semantic Resources	243
<i>Qiang Zhou (Tsinghua University, Beijing, China)</i>	

III Speech

Age Determination of Children in Preschool and Primary School Age with GMM-Based Supervectors and Support Vector Machines/Regression	253
<i>Tobias Bocklet, Andreas Maier, and Elmar Nöth (University of Erlangen-Nuremberg, Germany)</i>	
Language Acquisition: The Emergence of Words from Multimodal Input	261
<i>Louis ten Bosch and Lou Boves (Radboud University Nijmegen, Netherlands)</i>	
A Tagged Corpus-Based Study for Repeats and Self-Repairs Detection in French Transcribed Speech	269
<i>Rémi Bove (University of Provence, Aix-en-Provence, France)</i>	
Study on Speaker Adaptation Methods in the Broadcast News Transcription Task	277
<i>Petr Červa, Jindřich Ždánký, Jan Silovský, and Jan Nouza (Technical University of Liberec, Czech Republic)</i>	
Spoken Document Retrieval Based on Approximated Sequence Alignment	285
<i>Pere R. Comas and Jordi Turmo (Technical University of Catalonia, Barcelona, Spain)</i>	
Pitch Accents, Boundary Tones and Contours: Automatic Learning of Czech Intonation	293
<i>Tomáš Duběda and Jan Raab (Charles University in Prague, Czech Republic)</i>	
On the Use of MLP Features for Broadcast News Transcription	303
<i>Petr Fousek, Lori Lamel, and Jean-Luc Gauvain (LIMSI-CNRS, France)</i>	
Acoustic Modeling for Speech Recognition in Telephone Based Dialog System Using Limited Audio Resources	311
<i>Rok Gajšek, Janez Žibert, and France Mihelič (University of Ljubljana, Slovenia)</i>	
Performance Evaluation for Voice Conversion Systems	317
<i>Todor Ganchev, Alexandros Lazaridis, Iosif Mporas, and Nikos Fakotakis (University of Patras, Rion-Patras, Greece)</i>	
Influence of Reading Errors on the Text-Based Automatic Evaluation of Pathologic Voices	325
<i>Tino Haderlein, Elmar Nöth, Andreas Maier, Maria Schuster, and Frank Rosanowski (University of Erlangen-Nuremberg, Germany)</i>	

Czech Pitch Contour Modeling Using Linear Prediction	333
<i>Petr Horák (Academy of Sciences of the Czech Republic, Praha, Czech Republic)</i>	
Energy Normalization in Automatic Speech Recognition	341
<i>Nikša Jakovljević, Marko Janev, Darko Pekar, and Dragiša Mišković (University of Novi Sad, Serbia)</i>	
HMM-Based Speech Synthesis for the Greek Language	349
<i>Sotiris Karabetos, Pirros Tsiakoulis, Aimilios Chalamandaris, and Spyros Raptis (Institute for Language and Speech Processing, Athens, Greece)</i>	
Advances in Acoustic Modeling for the Recognition of Czech	357
<i>Jiří Kopecký, Ondřej Glembek, and Martin Karafiát (Brno University of Technology, Czech Republic)</i>	
Talking Head as Life Blog	365
<i>Ladislav Kunc (Czech Technical University in Prague, Czech Republic), Jan Kleindienst (IBM, Prague, Czech Republic), and Pavel Slavík (Czech Technical University in Prague, Czech Republic)</i>	
The Module of Morphological and Syntactic Analysis SMART	373
<i>Anastasia Leontyeva and Ildar Kagirov (St. Petersburg Institute for Informatics and Automation of RAS, Russia)</i>	
An Extension to the Sammon Mapping for the Robust Visualization of Speaker Dependencies	381
<i>Andreas Maier, Julian Exner, Stefan Steidl, Anton Batliner, Tino Haderlein, and Elmar Nöth (University of Erlangen-Nuremberg, Germany)</i>	
Analysis of Hypernasal Speech in Children with Cleft Lip and Palate	389
<i>Andreas Maier, Alexander Reuß, Christian Hacker, Maria Schuster, and Elmar Nöth (University of Erlangen-Nuremberg, Germany)</i>	
Syllable Based Language Model for Large Vocabulary Continuous Speech Recognition of Polish	397
<i>Piotr Majewski (University of Łódź, Poland)</i>	
Accent and Channel Adaptation for Use in a Telephone-Based Spoken Dialog System	403
<i>Kinfe Tadesse Mengistu and Andreas Wendemuth (Otto-von-Guericke University, Magdeburg, Germany)</i>	
Efficient Unit-Selection in Text-to-Speech Synthesis	411
<i>Aleš Mihelič and Jerneja Žganec Gros (Alpineon d.o.o., Ljubljana, Slovenia)</i>	

Prosodic Events Recognition in Evaluation of Speech-Synthesis System Performance	419
<i>France Mihelič, Boštjan Vesnicer, Janez Žibert (University of Ljubljana, Slovenia), and Elmar Nöth (University of Erlangen-Nuremberg, Germany)</i>	
UU Database: A Spoken Dialogue Corpus for Studies on Paralinguistic Information in Expressive Conversation	427
<i>Hiroki Mori, Tomoyuki Satake, Makoto Nakamura (Utsunomiya University, Japan), and Hideki Kasuya (International University of Health and Welfare, Japan)</i>	
Perceptually Motivated Sub-Band Decomposition for FDLF Audio Coding	435
<i>Petr Motlíček (IDIAP Research Institute, Martigny, Switzerland, Brno University of Technology, Czech Republic), Sriram Ganapathy (IDIAP Research Institute, Martigny, Switzerland, EPFL, Switzerland), Hynek Hermansky (IDIAP Research Institute, Martigny, Switzerland, Brno University of Technology, Czech Republic, EPFL, Switzerland), Harinath Garudadri (Qualcomm Inc., San Diego, United States), and Marios Athineos (International Computer Science Institute, Berkeley, United States)</i>	
Forced-Alignment and Edit-Distance Scoring for Vocabulary Tutoring Applications	443
<i>Serguei Pakhomov (University of Minnesota, United States), Jayson Richardson (University of North Carolina, United States), Matt Finholt-Daniel, and Gregory Sales (Seward Incorporated, United States)</i>	
Exploiting Contextual Information for Speech/Non-Speech Detection	451
<i>Sree Hari Krishnan Parthasarathi, Petr Motlíček, and Hynek Hermansky (IDIAP Research Institute, Martigny, Switzerland)</i>	
Identification of Speakers from Their Hum	461
<i>Hemant A. Patil, Robin Jain, and Prakhar Jain (Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India)</i>	
Reverse Correlation for Analyzing MLP Posterior Features in ASR	469
<i>Joel Pinto, Garimella S.V.S. Sivaram, and Hynek Hermansky (IDIAP Research Institute, Martigny, Swiss Federal Institute of Technology, Lausanne, Switzerland)</i>	
Acquisition of Telephone Data from Radio Broadcasts with Applications to Language Recognition	477
<i>Oldřich Plchot, Valiantsina Hubeika, Lukáš Burget, Petr Schwarz, and Pavel Matějka (Brno University of Technology, Czech Republic)</i>	

Multilingual Weighted Codebooks for Non-Native Speech Recognition	485
<i>Martin Raab, Rainer Gruhn (Harman Becker Automotive Systems, Ulm, Germany), and Elmar Nöth (University of Erlangen, Germany)</i>	
Prosodic Phrases and Semantic Accents in Speech Corpus for Czech TTS Synthesis	493
<i>Jan Romportl (University of West Bohemia, Czech Republic)</i>	
Making Speech Technologies Available in (Serviko) Romani Language	501
<i>Milan Rusko, Sakhia Darjaa, Marián Trnka (Institute of Informatics of the Slovak Academy of Sciences, Bratislava, Slovakia), Viliam Zeman (Office of the Plenipotentiary of the Government of the Slovak Republic for Romanies, Bratislava, Slovakia), and Juraj Glovňa (University of Constantinus, the Philosopher in Nitra, Slovakia)</i>	
Emulating Temporal Receptive Fields of Higher Level Auditory Neurons for ASR	509
<i>Garimella S.V.S. Sivaram and Hynek Hermansky (IDIAP Research Institute, Martigny, Swiss Federal Institute of Technology, Lausanne, Switzerland)</i>	
A Framework for Language-Independent Analysis and Prosodic Feature Annotation of Text Corpora	517
<i>Dimitris Spiliotopoulos, Georgios Petasis, and Georgios Kouroupetroglou (National and Kapodistrian University of Athens, Greece)</i>	
Quantification of Segmentation and F0 Errors and Their Effect on Emotion Recognition	525
<i>Stefan Steidl, Anton Batliner, Elmar Nöth, and Joachim Hornegger (University of Erlangen-Nuremberg, Germany)</i>	
Deep Syntactic Analysis and Rule Based Accentuation in Text-to-Speech Synthesis	535
<i>Antti Suni and Martti Vainio (University of Helsinki, Finland)</i>	
Error Prediction-Based Semi-automatic Segmentation of Speech Databases	543
<i>Marcin Szymański (Adam Mickiewicz University Foundation & Poznań University of Technology, Poland) and Stefan Grochowski (Poznań University of Technology, Poland)</i>	

IV Dialogue

Cognitive and Emotional Interaction	553
<i>Amel Achour, Jeanne Villaneau, and Dominique Duhaut (University of Southern Brittany, Lorient, France)</i>	

Architecture Model and Tools for Perceptual Dialog Systems	561
<i>Jan Cuřín and Jan Kleindienst (IBM, Prague, Czech Republic)</i>	
The Generation of Emotional Expressions for a Text-Based Dialogue Agent . . .	569
<i>Siska Fitriani (Delft University of Technology, Netherlands) and Leon J.M. Rothkrantz (Delft University of Technology & Netherlands Defence Academy, Netherlands)</i>	
Intelligent Voice Navigation of Spreadsheets: An Empirical Evaluation	577
<i>Derek Flood, Kevin Mc Daid, Fergal Mc Caffery, and Brian Bishop (Dundalk Institute of Technology, Ireland)</i>	
A Semi-automatic Wizard of Oz Technique for Let'sFly Spoken Dialogue System	585
<i>Alexey Karpov, Andrey Ronzhin, and Anastasia Leontyeva (St. Petersburg Institute for Informatics and Automation of RAS, Russia)</i>	
Developing a Dialogue System: How to Grant a Customer's Directive?	593
<i>Mare Koit, Olga Gerassimenko, Riina Kasterpalu, Andriela Rääbis, and Krista Strandson (University of Tartu, Estonia)</i>	
Dialogue-Based Processing of Graphics and Graphical Ontologies	601
<i>Ivan Kopeček and Radek Ošlejšek (Masaryk University, Brno, Czech Republic)</i>	
An Issue-Based Approach to Information Search Modelling: Analysis of a Human Dialog Corpus	609
<i>Alain Loisel, Jean-Philippe Kotowicz, and Nathalie Chaignaud (INSA of Rouen, Mont-Saint-Aignan, France)</i>	
Two-Level Fusion to Improve Emotion Classification in Spoken Dialogue Systems	617
<i>Ramón López-Cózar, Zoraida Callejas (University of Granada, Spain), Martin Kroul, Jan Nouza, and Jan Silovský (Technical University of Liberec, Czech Republic)</i>	
Automatic Semantic Annotation of Polish Dialogue Corpus	625
<i>Agnieszka Mykowiecka, Małgorzata Marciniak, and Katarzyna Głowińska (Polish Academy of Sciences, Warsaw, Poland)</i>	
Evaluation of the Slovak Spoken Dialogue System Based on ITU-T	633
<i>Stanislav Ondáš, Jozef Juhár, and Anton Čižmár (Technical University of Košice, Slovakia)</i>	

Shedding Light on a Troublesome Issue in NLIDBS: Word Economy in Query Formulation	641
<i>Rodolfo Pazos (CENIDET & ITCM, Mexico), René Santaolalaya S., Juan C. Rojas P., and Joaquín Pérez O. (CENIDET, Mexico)</i>	
Dialogue Based Text Editing	649
<i>Jaromír Plhák (Masaryk University, Brno, Czech Republic)</i>	
SSML for Arabic Language	657
<i>Noor Shaker, Mohamed Abou-Zleikha (University of Damascus, Syria), and Oumayma Al Dakkak (HIAST, Damascus, Syria)</i>	
Author Index	665