

# Contents

## Keynote Talk

A Glimpse Under the Surface: Language Understanding May Need Deep Syntactic Structure . . . . .	3
<i>Eva Hajičová</i>	

## Conference Papers

Robust Automatic Evaluation of Intelligibility in Voice Rehabilitation Using Prosodic Analysis . . . . .	11
<i>Tino Haderlein, Anne Schützenberger, Michael Döllinger, and Elmar Nöth</i>	
Personality-Dependent Referring Expression Generation. . . . .	20
<i>Ivandrė Paraboni, Danielle Sampaio Monteiro, and Alex Gwo Jen Lan</i>	
Big Five Personality Recognition from Multiple Text Genres . . . . .	29
<i>Vitor Garcia dos Santos, Ivandrė Paraboni, and Barbara Barbosa Claudino Silva</i>	
Automatic Classification of Types of Artefacts Arising During the Unit Selection Speech Synthesis . . . . .	38
<i>Jiří Přibil, Anna Přibilová, and Jindřich Matoušek</i>	
A Comparison of Lithuanian Morphological Analyzers . . . . .	47
<i>Jurgita Kapočiūtė-Dzikiėnė, Erika Rimkutė, and Loic Boizou</i>	
Constrained Deep Answer Sentence Selection. . . . .	57
<i>Ahmad Aghaebrahimian</i>	
Quora Question Answer Dataset . . . . .	66
<i>Ahmad Aghaebrahimian</i>	
Sentiment Analysis with Tree-Structured Gated Recurrent Units . . . . .	74
<i>Marcin Kuta, Mikołaj Morawiec, and Jacek Kitowski</i>	
Synthetic Speech in Therapy of Auditory Hallucinations . . . . .	83
<i>Kamil Sorokosz, Izabela Stefaniak, and Artur Janicki</i>	
Statistical Pronunciation Adaptation for Spontaneous Speech Synthesis . . . . .	92
<i>Raheel Qader, Gwénolé Lecorvé, Damien Lolive, Marie Tahon, and Pascale Sébillot</i>	

Machine Learning Approach to the Process of Question Generation . . . . .	102
<i>Miroslav Blšták and Viera Rozinajová</i>	
Automatic Extraction of Typological Linguistic Features from Descriptive Grammars . . . . .	111
<i>Shafqat Mumtaz Virk, Lars Borin, Anju Saxena, and Harald Hammarström</i>	
Text Punctuation: An Inter-annotator Agreement Study . . . . .	120
<i>Marek Boháč, Michal Rott, and Vojtěch Kovář</i>	
PDTSC 2.0 - Spoken Corpus with Rich Multi-layer Structural Annotation . . .	129
<i>Marie Mikulová, Jiří Mirovský, Anja Nedoluzhko, Petr Pajas, Jan Štěpánek, and Jan Hajič</i>	
Automatic Phonetic Segmentation Using the Kaldi Toolkit. . . . .	138
<i>Jindřich Matoušek and Michal Klíma</i>	
Language Independent Assessment of Motor Impairments of Patients with Parkinson's Disease Using i-Vectors. . . . .	147
<i>N. Garcia, J.C. Vásquez-Correa, J.R. Orozco-Arroyave, N. Dehak, and E. Nöth</i>	
ParCoLab: A Parallel Corpus for Serbian, French and English . . . . .	156
<i>Aleksandra Miletic, Dejan Stosic, and Saša Marjanović</i>	
Prosodic Phrase Boundary Classification Based on Czech Speech Corpora . . .	165
<i>Markéta Jůzová</i>	
Parliament Archives Used for Automatic Training of Multi-lingual Automatic Speech Recognition Systems. . . . .	174
<i>Jan Nouza and Radek Safarik</i>	
Recent Results in Speech Recognition for the Tatar Language . . . . .	183
<i>Aidar Khusainov</i>	
Data-Driven Identification of German Phrasal Compounds . . . . .	192
<i>Adrien Barbaresi and Katrin Hein</i>	
Automatic Preparation of Standard Arabic Phonetically Rich Written Corpora with Different Linguistic Units . . . . .	201
<i>Fadi Sindran, Firas Mualla, Tino Haderlein, Khaled Daqrouq, and Elmar Nöth</i>	
Adding Thesaurus Information into Probabilistic Topic Models. . . . .	210
<i>Natalia Loukachevitch and Michael Nokel</i>	

Dialogue Modelling in Multi-party Social Media Conversation . . . . .	219
<i>Subhabrata Dutta and Dipankar Das</i>	
Markov Text Generator for Basque Poetry . . . . .	228
<i>Aitzol Astigarraga, José María Martínez-Otzeta, Igor Rodriguez, Basilio Sierra, and Elena Lazkano</i>	
Neural Machine Translation for Morphologically Rich Languages with Improved Sub-word Units and Synthetic Data . . . . .	237
<i>Mārcis Pinnis, Rihards Krišlauks, Daiga Dekšne, and Toms Miks</i>	
Evaluation of Dictionary Creating Methods for Under-Resourced Languages . . . . .	246
<i>Eszter Simon and Iván Mittelholcz</i>	
Comparative Evaluation and Integration of Collocation Extraction Metrics . . .	255
<i>Victor Zakharov</i>	
Errors in Inflection in Czech as a Second Language and Their Automatic Classification . . . . .	263
<i>Tomáš Jelínek</i>	
Speaker Model to Monitor the Neurological State and the Dysarthria Level of Patients with Parkinson's Disease. . . . .	272
<i>J.C. Vásquez-Correa, R. Castrillón, T. Arias-Vergara, J.R. Orozco-Arroyave, and E. Nöth</i>	
A Lightweight Regression Method to Infer Psycholinguistic Properties for Brazilian Portuguese. . . . .	281
<i>Leandro Borges dos Santos, Magali Sanches Duran, Nathan Siegle Hartmann, Arnaldo Candido Jr., Gustavo Henrique Paetzold, and Sandra Maria Aluisio</i>	
Open-Domain Non-factoid Question Answering . . . . .	290
<i>Maria Khvalchik and Anagha Kulkarni</i>	
Linguistic Features as Evidence for Historical Context Interpretation . . . . .	299
<i>Jyi-Shane Liu, Ching-Ying Lee, and Hua-Yuan Hsueh</i>	
Morphosyntactic Annotation of Historical Texts. The Making of the Baroque Corpus of Polish. . . . .	308
<i>Witold Kieraś, Dorota Komosińska, Emanuel Modrzejewski, and Marcin Woliński</i>	
Last Syllable Unit Penalization in Unit Selection TTS . . . . .	317
<i>Markéta Jůzová, Daniel Tihelka, and Radek Skarnitzl</i>	

On Multilingual Training of Neural Dependency Parsers . . . . .	326
<i>Michał Zapotoczny, Paweł Rychlikowski, and Jan Chorowski</i>	
Meaning Extensions, Word Component Structures and Their Distribution: Linguistic Usages Containing Body-Part Terms Liǎn/Miàn, Yǎn/Mù and Zuǐ/Kǒu in Taiwan Mandarin . . . . .	335
<i>Hsiao-Ling Hsu, Huei-ling Lai, and Jyi-Shane Liu</i>	
An Unification-Based Model for Attitude Prediction . . . . .	344
<i>Manfred Klenner</i>	
Optimal Number of States in HMM-Based Speech Synthesis . . . . .	353
<i>Zdeněk Hanzlíček</i>	
Temporal Feature Space for Text Classification. . . . .	362
<i>Stefano Giovanni Rizzo and Danilo Montesi</i>	
Parkinson's Disease Progression Assessment from Speech Using a Mobile Device-Based Application . . . . .	371
<i>T. Arias-Vergara, P. Klumpp, J.C. Vásquez-Correa, J.R. Orozco-Arroyave, and E. Nöth</i>	
A New Corpus of Collaborative Dialogue Produced Under Cognitive Load Using a Driving Simulator . . . . .	380
<i>George Christodoulides</i>	
Phonetic Segmentation Using Knowledge from Visual and Perceptual Domain . . . . .	393
<i>Bhavik Vachhani, Chitrlekha Bhat, and Sunil Koppurapu</i>	
The Impact of Inaccurate Phonetic Annotations on Speech Recognition Performance . . . . .	402
<i>Radek Safarik and Lukas Mateju</i>	
Automatic Detection of Parkinson's Disease: An Experimental Analysis of Common Speech Production Tasks Used for Diagnosis . . . . .	411
<i>Anna Pompili, Alberto Abad, Paolo Romano, Isabel P. Martins, Rita Cardoso, Helena Santos, Joana Carvalho, Isabel Guimarães, and Joaquim J. Ferreira</i>	
Unified Simplified Grapheme Acoustic Modeling for Medieval Latin LVCSR . . . . .	420
<i>Lili Szabó, Péter Mihajlik, András Balog, and Tibor Fegyó</i>	
Experiments with Segmentation in an Online Speaker Diarization System . . .	429
<i>Marie Kunešová, Zbyněk Zajíc, and Vlasta Radová</i>	

Spatiotemporal Convolutional Features for Lipreading . . . . .	438
<i>Karel Paleček</i>	
Could Emotions Be Beneficial for Interaction Quality Modelling in Human-Human Conversations? . . . . .	447
<i>Anastasiia Spirina, Wolfgang Minker, and Maxim Sidorov</i>	
Multipoint Neighbor Embedding . . . . .	456
<i>Adrian Lancucki and Jan Chorowski</i>	
Significance of Interaction Parameter Levels in Interaction Quality Modelling for Human-Human Conversation . . . . .	465
<i>Anastasiia Spirina, Alina Skorokhod, Tatiana Karaseva, Iana Polonskaia, and Maxim Sidorov</i>	
Ship-LemmaTagger: Building an NLP Toolkit for a Peruvian Native Language . . . . .	473
<i>José Pereira-Noriega, Rodolfo Mercado-Gonzales, Andrés Melgar, Marco Sobrevilla-Cabezudo, and Arturo Oncevay-Marcos</i>	
A Study of Abstractive Summarization Using Semantic Representations and Discourse Level Information. . . . .	482
<i>Gregory César Valderrama Vilca and Marco Antonio Sobrevilla Cabezudo</i>	
Development and Integration of Natural Brazilian Portuguese Synthetic Voices to Framework FIVE . . . . .	491
<i>Danilo S. Barbosa, Byron L.D. Bezerra, and Alexandre M.A. Maciel</i>	
Fine-Tuning Word Embeddings for Aspect-Based Sentiment Analysis . . . . .	500
<i>Duc-Hong Pham, Thi-Thanh-Tan Nguyen, and Anh-Cuong Le</i>	
Recognition of the Electrolaryngeal Speech: Comparison Between Human and Machine . . . . .	509
<i>Petr Stanislav, Josef V. Psutka, and Josef Psutka</i>	
<b>Author Index . . . . .</b>	<b>519</b>

Text, Speech, and Dialogue

20th International Conference, TSD 2017, Prague,  
Czech Republic, August 27-31, 2017, Proceedings

Ekstein, K.; Matoušek, V. (Eds.)

2017, XXI, 520 p. 94 illus., Softcover

ISBN: 978-3-319-64205-5