

# Contents

## Text

Grammatical Annotation of Historical Portuguese: Generating a Corpus-Based Diachronic Dictionary. . . . .	3
<i>Eckhard Bick and Marcos Zampieri</i>	
Generating of Events Dictionaries from Polish WordNet for the Recognition of Events in Polish Documents. . . . .	12
<i>Jan Kocoń and Michał Marcińczuk</i>	
Building Corpora for Stylometric Research. . . . .	20
<i>Jan Švec and Jan Rygl</i>	
Constraint-Based Open-Domain Question Answering Using Knowledge Graph Search . . . . .	28
<i>Ahmad Aghaebrahimian and Filip Jurčiček</i>	
A Sentiment-Aware Topic Model for Extracting Failures from Product Reviews . . . . .	37
<i>Elena Tutubalina</i>	
Digging Language Model – Maximum Entropy Phrase Extraction. . . . .	46
<i>Jakub Kanis</i>	
Vive la Petite Différence! Exploiting Small Differences for Gender Attribution of Short Texts . . . . .	54
<i>Filip Graliński, Rafał Jaworski, Łukasz Borchmann, and Piotr Wierzchoń</i>	
Towards It-CMC: A Fine-Grained POS Tagset for Italian Linguistic Analysis . . . . .	62
<i>Claudio Russo</i>	
FAQIR – A Frequently Asked Questions Retrieval Test Collection . . . . .	74
<i>Mladen Karan and Jan Šnajder</i>	
Combining Dependency Parsers Using Error Rates . . . . .	82
<i>Tomáš Jelínek</i>	
A Modular Chain of NLP Tools for Basque . . . . .	93
<i>Arantxa Otegi, Nerea Ezeiza, Iakes Goenaga, and Gorka Labaka</i>	

Speech-to-Text Summarization Using Automatic Phrase Extraction from Recognized Text . . . . .	101
<i>Michal Rott and Petr Červa</i>	
Homonymy and Polysemy in the Czech Morphological Dictionary . . . . .	109
<i>Jaroslava Hlaváčová</i>	
Cross-Language Dependency Parsing Using Part-of-Speech Patterns . . . . .	117
<i>Peter Bednár</i>	
Assessing Context for Extraction of Near Synonyms from Product Reviews in Spanish . . . . .	125
<i>Sofía N. Galicia-Haro and Alexander F. Gelbukh</i>	
Gathering Information About Word Similarity from Neighbor Sentences . . . .	134
<i>Natalia Loukachevitch and Aleksei Alekseev</i>	
Short Messages Spam Filtering Using Sentiment Analysis . . . . .	142
<i>Enaitz Ezpeleta, Urko Zurutuza, and José María Gómez Hidalgo</i>	
Preliminary Study on Automatic Recognition of Spatial Expressions in Polish Texts . . . . .	154
<i>Michał Marcińczuk, Marcin Oleksy, and Jan Wieczorek</i>	
Topic Modeling over Text Streams from Social Media . . . . .	163
<i>Miroslav Smatana, Ján Paralič, and Peter Butka</i>	
Neural Networks for Featureless Named Entity Recognition in Czech . . . . .	173
<i>Jana Straková, Milan Straka, and Jan Hajič</i>	
SubGram: Extending Skip-Gram Word Representation with Substrings . . . . .	182
<i>Tom Kocmi and Ondřej Bojar</i>	
WordSim353 for Czech . . . . .	190
<i>Silvie Cinková</i>	
Automatic Restoration of Diacritics for Igbo Language . . . . .	198
<i>Ignatius Ezeani, Mark Hepple, and Ikechukwu Onyenwe</i>	
Predicting Morphologically-Complex Unknown Words in Igbo. . . . .	206
<i>Ikechukwu E. Onyenwe and Mark Hepple</i>	
Morphosyntactic Analyzer for the Tibetan Language: Aspects of Structural Ambiguity . . . . .	215
<i>Alexei Dobrov, Anastasia Dobrova, Pavel Grokhovskiy, Nikolay Soms, and Victor Zakharov</i>	
Automatic Question Generation Based on Analysis of Sentence Structure . . .	223
<i>Miroslav Blšták and Viera Rozinajová</i>	

CzEng 1.6: Enlarged Czech-English Parallel Corpus with Processing Tools Dockered . . . . .	231
<i>Ondřej Bojar, Ondřej Dušek, Tom Kocmi, Jindřich Libovický, Michal Novák, Martin Popel, Roman Sudarikov, and Dušan Variš</i>	
Using Alliteration in Authorship Attribution of Historical Texts . . . . .	239
<i>Lubomir Ivanov</i>	
Collecting Facebook Posts and WhatsApp Chats: Corpus Compilation of Private Social Media Messages . . . . .	249
<i>Lieke Verheijen and Wessel Stoop</i>	
A Dynamic Programming Approach to Improving Translation Memory Matching and Retrieval Using Paraphrases . . . . .	259
<i>Rohit Gupta, Constantin Orăsan, Qun Liu, and Ruslan Mitkov</i>	
AQA: Automatic Question Answering System for Czech . . . . .	270
<i>Marek Medved' and Aleš Horák</i>	
Annotation of Czech Texts with Language Mixing . . . . .	279
<i>Zuzana Nevěřilová</i>	
Evaluation and Improvements in Punctuation Detection for Czech. . . . .	287
<i>Vojtěch Kovář, Jakub Machura, Kristýna Zemková, and Michal Rott</i>	
Annotated Amharic Corpora . . . . .	295
<i>Pavel Rychlý and Vít Suchomel</i>	

## Speech

Evaluation of TTS Personification by GMM-Based Speaker Gender and Age Classifier. . . . .	305
<i>Jiří Přibíl, Anna Přibílová, and Jindřich Matoušek</i>	
Grapheme to Phoneme Translation Using Conditional Random Fields with Re-Ranking. . . . .	314
<i>Stephen Ash and David Lin</i>	
On the Influence of the Number of Anomalous and Normal Examples in Anomaly-Based Annotation Errors Detection . . . . .	326
<i>Jindřich Matoušek and Daniel Tihelka</i>	
Unit-Selection Speech Synthesis Adjustments for Audiobook-Based Voices . . .	335
<i>Jakub Vít and Jindřich Matoušek</i>	
The Custom Decay Language Model for Long Range Dependencies . . . . .	343
<i>Mittul Singh, Clayton Greenberg, and Dietrich Klakow</i>	

Voice Activity Detector (VAD) Based on Long-Term Mel Frequency Band Features. . . . .	352
<i>Sergey Salishev, Andrey Barabanov, Daniil Kocharov, Pavel Skrelin, and Mikhail Moiseev</i>	
Difficulties with Wh-Questions in Czech TTS System . . . . .	359
<i>Markéta Jůzová and Daniel Tihelka</i>	
Tools rPraat and mPraat: Interfacing Phonetic Analyses with Signal Processing. . . . .	367
<i>Tomáš Bořil and Radek Skarnitzl</i>	
A Composition Algorithm of Compact Finite-State Super Transducers for Grapheme-to-Phoneme Conversion. . . . .	375
<i>Žiga Golob, Jerneja Žganec Gros, Vitomir Štruc, France Mihelič, and Simon Dobrišek</i>	
Embedded Learning Segmentation Approach for Arabic Speech Recognition . . . . .	383
<i>Hamza Frihia and Halima Bahi</i>	
KALDI Recipes for the Czech Speech Recognition Under Various Conditions . . . . .	391
<i>Petr Mizera, Jiří Fiala, Aleš Brich, and Petr Pollak</i>	
Glottal Flow Patterns Analyses for Parkinson’s Disease Detection: Acoustic and Nonlinear Approaches . . . . .	400
<i>Elkyn Alexander Belalcázar-Bolaños, Juan Rafael Orozco-Arroyave, Jesús Francisco Vargas-Bonilla, Tino Haderlein, and Elmar Nöth</i>	
Correction of Prosodic Phrases in Large Speech Corpora . . . . .	408
<i>Zdeněk Hanzlíček</i>	
Relevant Documents Selection for Blind Relevance Feedback in Speech Information Retrieval. . . . .	418
<i>Lucie Skorkovská</i>	
Investigation of Bottle-Neck Features for Emotion Recognition. . . . .	426
<i>Anna Popková, Filip Povolný, Pavel Matějka, Ondřej Glembek, František Grézl, and Jan “Honza” Černocký</i>	
Classification of Speaker Intoxication Using a Bidirectional Recurrent Neural Network . . . . .	435
<i>Kim Berninger, Jannis Hoppe, and Benjamin Milde</i>	
Training Maxout Neural Networks for Speech Recognition Tasks . . . . .	443
<i>Aleksey Prudnikov and Maxim Korenevsky</i>	

An Efficient Method for Vocabulary Addition to WFST Graphs . . . . .	452
<i>Anna Bulusheva, Alexander Zatvornitskiy, and Maxim Korenevsky</i>	

## Dialogue

Influence of Reverberation on Automatic Evaluation of Intelligibility with Prosodic Features. . . . .	461
<i>Tino Haderlein, Michael Döllinger, Anne Schützenberger, and Elmar Nöth</i>	
Automatic Scoring of a Sentence Repetition Task from Voice Recordings . . .	470
<i>Meysam Asgari, Allison Sliter, and Jan Van Santen</i>	
Platon: Dialog Management and Rapid Prototyping for Multilingual Multi-user Dialog Systems . . . . .	478
<i>Martin Gropp, Anna Schmidt, Thomas Kleinbauer, and Dietrich Klakow</i>	
How to Add Word Classes to the Kaldi Speech Recognition Toolkit . . . . .	486
<i>Axel Horndasch, Caroline Kaufhold, and Elmar Nöth</i>	
Starting a Conversation: Indexical Rhythmical Features Across Age and Gender (A Corpus Study) . . . . .	495
<i>Tatiana Sokoreva and Tatiana Shevchenko</i>	
Classification of Utterance Acceptability Based on BLEU Scores for Dialogue-Based CALL Systems. . . . .	506
<i>Reiko Kuwa, Xiaoyun Wang, Tsuneo Kato, and Seiichi Yamamoto</i>	
A Unified Parser for Developing Indian Language Text to Speech Synthesizers. . . . .	514
<i>Arun Baby, Nishanthi N.L., Anju Leela Thomas, and Hema A. Murthy</i>	
Influence of Expressive Speech on ASR Performances: Application to Elderly Assistance in Smart Home. . . . .	522
<i>Frédéric Aman, Véronique Aubergé, and Michel Vacher</i>	
From Dialogue Corpora to Dialogue Systems: Generating a Chatbot with Teenager Personality for Preventing Cyber-Pedophilia . . . . .	531
<i>Ángel Callejas-Rodríguez, Esaú Villatoro-Tello, Ivan Meza, and Gabriela Ramírez-de-la-Rosa</i>	
Automatic Syllabification and Syllable Timing of Automatically Recognized Speech – for Czech . . . . .	540
<i>Marek Boháč, Lukáš Matějů, Michal Rott, and Radek Šafařík</i>	
<b>Author Index</b> . . . . .	549

Text, Speech, and Dialogue

19th International Conference, TSD 2016, Brno , Czech  
Republic, September 12-16, 2016, Proceedings

Sojka, P.; Horák, A.; Kopeček, I.; Pala, K. (Eds.)

2016, XXI, 550 p. 109 illus., Softcover

ISBN: 978-3-319-45509-9