# Contents

## Speech

Improved Syllable-Based Text to Speech Synthesis for Tone Language Systems ............................................................... 3  
*Moses Ekpenyong, EmemObong Udoh, Escor Udosen, and Eno-Abasi Urua*

Slovak Automatic Dictation System for Judicial Domain ......................... 16  
*Milan Rusko, Jozef Juhár, Marián Trnka, Ján Staš, Sakhia Darjaa, Daniel Hládek, Róbert Sabo, Matúš Pleva, Marián Ritomský, and Martin Lojka*

The Intonation of Backchannel Tokens in Italian Collaborative Dialogues ... 28  
*Michelina Savino*

## Parsing

Improved Parsing for Arabic by Combining Diverse Dependency Parsers .... 43  
*Maytham Alabbas and Allan Ramsay*

Parsing Coordination Extragrammatically ........................................ 55  
*Valmi Dufour-Lussier, Bruno Guillaume, and Guy Perrier*

Lexical Disambiguation in LTAG Using Left Context .......................... 67  
*Claire Gardent, Yannick Parmentier, Guy Perrier, and Sylvain Schmitz*

## Computational Semantics

Resolving Anaphors in Sanskrit ................................................. 83  
*Madhav Gopal and Girish Nath Jha*

Unsupervised Coreference Resolution Using a Graph Labeling Approach ... 93  
*Nafise Sadat Moosavi and GholamReza GhassemSani*

The XCDC Relations as a Spatio-Temporal Ontology ......................... 104  
*Jędrzej Osiński*

Normalization of Term Weighting Scheme for Sentiment Analysis ........ 116  
*Alexander Pak, Patrick Paroubek, Amel Fraisse, and Gil Francopoulo*

Thel, a Language for Utterance Generation in the Thetos System .......... 129  
*Julia Romaniuk, Nina Suszczańska, and Przemysław Szmal*
An Algorithm for Inconsistency Management in Spatial Knowledge Integration  ........................................ 141
Marcin Walas and Krzysztof Jassem

Text Analysis

Information Extraction for Czech Based on Syntactic Analysis ........... 155
Vít Baisa and Vojtěch Kovář

Applying Rule-Based Normalization to Different Types of Historical Texts — An Evaluation ................................. 166
Marcel Bollmann, Florian Petran, and Stefanie Dipper

A Rule Based Method for the Identification of TAM Features in a PoS Tagged Corpus ........................................... 178
Narayan Choudhary, Pramod Pandey, and Girish Nath Jha

Digging for Names in the Mountains: Combined Person Name Recognition and Reference Resolution for German Alpine Texts ......................... 189
Sarah Ebling, Rico Sennrich, and David Klaper

Enhancing Labeled Data Using Unlabeled Data for Topic Tracking ........ 201
Fumiyo Fukumoto, Yoshimi Suzuki, and Takeshi Yamamoto

Temporal Expression Recognition Using Dependency Trees .............. 213
Paweł Mazur and Robert Dale

Pattern Mining for Named Entity Recognition ................................ 226
Damien Nouvel, Jean-Yves Antoine, and Nathalie Friburger

Lexical Bundles in Swedish Secondary School Textbooks .................. 238
Judy Ribeck and Lars Borin

Text Annotation

Active Learning to Speed-Up the Training Process for Dialogue Act Labelling ....................................................... 253
Fabrizio Ghigi, Carlos-D. Martínez-Hinarejos, and José-Miguel Benedí

Direct and Wordgraph-Based Confidence Measures in Dialogue Annotation with N-Gram Transducers ............................... 264
Carlos-D. Martínez-Hinarejos, Vicent Tamarit, and José-Miguel Benedí

Orwell’s 1984 — From Simple to Multi-word Units .......................... 276
Cvetana Krstev, Duško Vitas, and Aleksandra Trtovac
Application of Audio and Video Processing Methods for Language Research and Documentation: The AVATecH Project ...................... 288
  Przemyslaw Lenkiewicz, Sebastian Drude, Anna Lenkiewicz, Binyam Gebrekidan Gebre, Stefano Masneri, Oliver Schreer, Jochen Schwenninger, and Rolf Bardeli

Language Resources: General Issues

Crowdsourcing for Language Resource Development: Criticisms About Amazon Mechanical Turk Overpowering Use ......................... 303
  Karén Fort, Gilles Adda, Benoît Sagot, Joseph Mariani, and Alain Couillault

Extending a Tool Resource Framework with U-Compare .................... 315
  Michael Rosner, Andrew Attard, Paul Thompson, Albert Gatt, and Sophia Ananiadou

Language Resources: Ontologies and Wordnets

Aligning GermaNet Senses with Wiktionary Sense Definitions .............. 329
  Verena Henrich, Erhard Hinrichs, and Tatiana Vodolazova

A Tool for Transforming WordNet-Like Databases .......................... 343
  Marek Kubis

KABA Subject Heading Language as the Main Resource Subject
Organization Tool in a Semantic Knowledge Base .......................... 356
  Cezary Mazurek, Krzysztof Sielski, Justyna Walkowska, and Marcin Werla

Enhancing Tagging Systems by Wordnet Based Ontologies ................ 367
  Jacek Marciniak

Natural Language Ontology of Action: A Gap with Huge Consequences
for Natural Language Understanding and Machine Translation ............ 379
  Massimo Moneglia

Classification-Based Extension of Wordnets from Heterogeneous Resources ... 396
  Benoît Sagot and Darja Fišer

PolNet – Polish WordNet .................................................. 408
  Zygmunt Vetulani

Machine Translation

Improving the Distribution of N-Grams in Phrase Tables Obtained
by the Sampling-Based Method .............................................. 419
  Juan Luo, Adrien Lardilleux, and Yves Lepage
Marker-Based Chunking in Eleven European Languages
for Analogy-Based Translation ............................................. 432
Kota Takeya and Yves Lepage

Comparing CBMT Approaches for German-Romanian ............... 445
Monica Gavrila and Natalia Eliia

Text Genre – An Unexplored Parameter in Statistical Machine Translation .... 456
Monica Gavrila and Cristina Vertan

Problems Concerning Less Resourced Languages

Detecting Gaps in Language Resources and Tools in the Project CESAR ... 471
Marko Tadić, Tamás Váradi, Radovan Garabík, Svetla Koeva,
Maciej Ogrodniczuk, and Duško Vitas

A First LVCSR System for Luxembourgish, a Low-Resourced European
Language ................................................................. 479
Martine Adda-Decker, Lori Lamel, Gilles Adda, and Thomas Lavergne

Developing LRs for Non-scheduled Indian Languages ................. 491
Ritesh Kumar, Bornini Lahiri, and Deepak Alok

Quizzes on Tap: Exporting a Test Generation System from One
Less-Resourced Language to Another ..................................... 502
Montse Maritxalar, Elaine Uí Donnchadha, Jennifer Foster,
and Monica Ward

A Multilingual Text Normalization Approach ............................ 515
Brigitte Bigi

Creating Multilingual Parallel Corpora in Indian Languages ............ 527
Narayan Choudhary and Girish Nath Jha

Inducing Grammars from IGT ............................................. 538
Lars Hellan and Dorothee Beermann

Author Index .............................................................. 549
Human Language Technology Challenges for Computer Science and Linguistics
5th Language and Technology Conference, LTC 2011, Poznań, Poland, November 25--27, 2011, Revised Selected Papers
Vetulani, Z.; Mariani, J. (Eds.)
2014, XVI, 550 p. 109 illus., Softcover
ISBN: 978-3-319-08957-7