

Contents

List of Reviewers	6
Preface by Zygmunt Vetulani and Patrick Paroubek	7
Nilufar Abdurakhmonova, Alisher Ismailov and Rano Sayfullaeva / Turkic language stemmer python package for Natural Language Processing	9
Jolanta Bachan, Marek Kubis, Natalia Maria Łozińska and Marta Kunegunda Witkowska / Towards analysis of hegemonic masculinity in the dialogues of Polish novels	14
Françoise Bacquellaine / DeepL and Google Translate Translating Portuguese Multi-Word Units into French: Progress, Decline and Remaining Challenges (2019-2023)	19
Brigitte Bigi / An analysis of produced versus predicted French Cued Speech keys	24
Lynne Bowker / Combining plain language and machine translation for science communication	29
Alexandra Ciobotaru, Diana Constantina Hoefels and Stefan Daniel Dumitrescu / Emotion Signals for Sexist and Offensive Language Detection: A Multi-task Learning Approach	34
Grażyna Demenko, Paweł Skórzewski, Tomasz Kuczmarski and Mikołaj Pieniowski / Linguistic Information Extraction from Text-based Web to Discover Criminal Activity	39
Matthew Eget, Xuchen Yang and Yves Lepage / A Study in the Generation of Multilingually Parallel Middle Sentences	45
Iris Eshkol-Taravella, Angèle Barbedette, Xingyu Liu and Valentin-Gabriel Soumah / Automatic Classification of Spontaneous vs Prepared Questions in Speech Transcriptions	50
Bojan Evkoski and Senja Pollak / XAI in Computational Linguistics - Understanding Political Leanings in the Slovenian Parliament	56
Swapnil Fadte, Edna Vaz, Atul Kr Ojha, Ramdas Karmali and Jyoti Pawar / Empirical Analysis of Oral and Nasal Vowels of Konkani	62
Rashel Fam and Yves Lepage / Investigating parallelograms: Assessing several word embedding spaces against various analogy test sets in several languages using approximation	68
Ewelina Gajewska and Barbara Konat / Text Classification for Subjective Phenomena on Disaggregated Data and Rater Behaviour	73
Thomas Gerald, Sofiane Ettayeb, Louis Tamames, Ha Quang Le, Patrick Paroubek and Anne Vilnat / A new approach to generate teacher-like questions guided by text spans extraction	78
Monika Grajzer, Mikołaj Pabiszczak, Agnieszka Bętkowska Cavalcante and Michał Raszewski / A Voice-Based Neural Network System for Accessing Embedded Home Automation Devices	85
Ryo Hashimoto, Masashi Takeshita, Rafal Rzepka and Kenji Araki / Development of Japanese WSC273 Wino-grad Schema Challenge Dataset and Comparison between Japanese and English BERT Baselines	91
Friederike Hohl and Bettina Braun / Using amplitude envelope modulation spectra to capture differences between rhetorical and information-seeking questions	96
Irina Illina and Dominique Fohr / Semantic Information Investigation for Transformer-based Rescoring of N-best Speech Recognition	101
Balázs Indig and Dániel Lévai / I'm Smarter than the Average BERT! – Testing Language Models Against Humans in a Word Guessing Game	106
Balázs Indig and Tímea Borbála Bajzát / Bags and Mosaics: Semi-automatic Identification of Auxiliary Verbal Constructions for Agglutinative Languages	111
Patrick Juola / Stylometry: a Need for Standard	117

Olha Kanishcheva / ASSETUK: a Dataset for Ukrainian Text Simplification	122
Oleg Kapanadze, Nunu Kapanadze, Gideon Kotzé and Natia Putkaradze / Unsupervised Syntactic Analysis of the Georgian Language Clause	126
Irakli Kardava / Optimizations of Some Well-Known NLP Algorithms	131
Maciej Karpinski, Ewa Jarmolowicz-Nowikow, Katarzyna Klessa, Janusz Taborek and Michał Piosik / DARIAH.PL MultiCo Multimodal Corpus	136
Elmurod Kuriyozov, Ulugbek Salaev, Sanatbek Matlatipov and Gayrat Matlatipov / Text classification dataset and analysis for Uzbek language	141
Tiziano Labruna and Bernardo Magnini / Simulating Domain Changes in Task-oriented Dialogues	146
Koena Ronny Mabokela, Tim Schlippe, Mpho Raborife and Turgay Celik / Language-Independent Sentiment Labelling with Distant Supervision: A Case Study for English, Sepedi and Setswana	152
Khabibulla Madatov, Sanatbek Matlatipov and Mersaid Aripov / Uzbek text's correspondence with the educational potential of pupils: a case study of the School corpus	156
Khabibulla Madatov, Shukurla Bekchanov and Jernej Vičič / Uzbek text summarization based on TF-IDF	161
Margot Madina, Itziar Gonzalez-Dios and Melanie Siegel / Easy-to-Read in Germany: a Survey on its Current State and Available Resources	166
Tanjim Mahmud, Michal Ptaszynski and Fumito Masui / Vulgar Remarks Detection in Chittagonian Dialect of Bangla	171
Madina Mansurova, Nurgali Kadyrbek and Talshyn Sarsembayeva / Aspect Based Sentiment Analysis by Morphological Features of The Kazakh Language	176
Hansjörg Mixdorff and Roberto Togneri / Evaluation of Foreign Accent Prosody in L2 English Using CNNs	181
Emese K Molnár and Andrea Dömötör / Experiments on error detection in morphological annotation	186
Jürgen Neyer, Sassan Gholiagha and Mitja Sienknecht / Do Arguments Migrate? Using NLP for Understanding Academia	191
Ata Nizamoglu, Lea Dahm, Talia Sari, Vera Schmitt, Salar Mohtaj and Sebastian Möller / A Transfer Learning Approach for SDGs Classification of Sustainability Reports	196
Nor Saiful Azam Bin Nor Azmi, Michal Ptaszynski, Juuso Eronen, Karol Nowakowski and Fumito Masui / Token and Part-of-Speech Fusion for Pretraining of Transformers with Application in Automatic Cyberbullying Detection	201
Anouar Nouri, Salar Mohtaj, Sebastian Möller and Tilman Lesch / A Comparative Study of Claim Extraction Techniques Leveraging Transformer-based Pre-Trained Models	206
Jędrzej Osiński and Daniel Rimoli / User Experience aspects of the WordNet-based digital asset search enhancement	213
Moumita Pakrashi, Brigitte Bigi and Shakuntala Mahanta / Resources Creation of Bengali for SPPAS	218
Marco Palomino, Rohan Allen and Aditya Padmanabhan Varma / Depression in the Times of COVID-19: A Machine Learning Analysis Based on the Profile of Mood States	223
Paweł Paściak, Danijel Koržinek and Dariusz Czerski / Self-supervised Domain Adaptation of Statistical Language Models for Automatic Speech Recognition	229
Mariia Razno / Comparative Analysis Of Speech-To-Text Systems For Ukrainian Dialects	234
Aleksandra Rewerska and Antonina Świdurska / RPGs: small-scale register analysis using the taxonomy of discourse units	240
Aleksandra Rykowska and Konrad Juszczyk / Syllab: program for automatic sentiment analysis of poetry based on frequencies of phonetic units	245
Rafał Rzepka, Shinji Muraji and Akihiko Obayashi / Utilizing Wikipedia for Retrieving Synonyms of Trade Security-related Technical Terms	250
Felix Schneider, Sven Sickert, Phillip Brandes, Sophie Marshall and Joachim Denzler / Hard is the Task, the Samples are Few: A German Chiasmus Dataset	255

Emre Sevindik, Elif Ergin, Kübra Erat, Pınar Onay Durdu / Artificial Neural Networks based Baby Sign Language Recognition via Wearable Sensors	261
Maksud Sharipov, Elmurod Kuriyozov, Ollabergan Yuldashev and Og‘abek Sobirov / UzbekTagger: The rule-based POS tagger for Uzbek language	267
Nomsa Skosana, Respect Mlambo and Muzi Matfunjwa / Strategies for creating corpora and language resources for under-resourced South African indigenous languages	272
Joanna Szwoch, Mateusz Staszko, Rafal Rzepka and Kenji Araki / Sentiment Analysis of Polish Online News Covering Controversial Topics – Comparison Between Lexicon and Statistical Approaches	277
Ilija Tavchioski, Marko Robnik-Šikonja and Senja Pollak / Detection of depression on social networks using transformers and ensembles	282
Irina Temnikova, Silvia Gargova, Ruslana Margova, Veneta Kireva, Ivo Dzumerov, Tsvetelina Stefanova and Hristiana Krasteva / New Bulgarian Resources for Studying Deception and Detecting Disinformation	288
Gaurish Thakkar, Nives Mikelić Preradović and Marko Tadić / CroSentiNews 2.0: A Sentence-Level News Sentiment Corpus	294
Ualsher Tukeyev, Gulstan Akhmet, Nargiza Gabdullina, Aliya Turganbayeva and Tolganai Balabekova / Kazakh-Uzbek Machine Translation on the Base of Complete Set of Endings Model	299
Sireesha Vakada, Anudeep Chaluvadi, Mounika Marreddy and Radhika Mamidi / IndicSumm: Summarization Resource Creation for Eight Indian Languages	304
Zygmunt Vetulani and Peter Odrakiewicz / Challenges and a New Paradigm Frontier of Human Language Technology Applications in Business Management, Business Communication in Organizations and Society	309
Lu Wang, Michał Ptaszynski, Paweł Dybala, Yuki Urabe, Rafal Rzepka and Fumito Masui / Improving Performance of Affect Analysis System by Expanding Affect Lexicon	314
Yaling Wang, Bartholomäus Wloka and Yves Lepage / Translation Memory Principle in Neural Machine Translation: A Multilingual and Multidirectional Comparison	320
Liyan Wang, Zhicheng Pan, Haotong Wang, Xinbo Zhao and Yves Lepage / Solving Sentence Analogies by Using Embedding Spaces Combined with a Vector-to-Sequence Decoder or by Fine-Tuning Pre-trained Language Models	325
Yizhe Wang, Béatrice Daille and Nabil Hathout / Exploring synonymy relation between multi-word terms in distributional semantic models	331
Liang Xu, Elaine Ui Dhonnchadha and Monica Ward / Exploring the Synergies between Technology and Socio-Cultural Approaches in CALL for Less Commonly Taught languages	337
Nicolas Zampieri, Irina Illina and Dominique Fohr / Improving Hate Speech Detection with Self-Attention Mechanism and Multi-task Learning	343
Yiyang Zhang, Masashi Takeshita, Rafal Rzepka and Kenji Araki / Utilizing BERT with Auxiliary Sentences Generation to Improve Accuracy of Japanese Aspect-based Sentiment Analysis Task	348
Andrzej Zydrón, Rafał Jaworski and Szymon Kaczmarek / Large Language Models and the future of the Localization Industry	353
Author index	357