

**INTERNATIONAL CONFERENCE
RECENT ADVANCES
IN NATURAL LANGUAGE PROCESSING**

RANLP 2019

**Natural Language Processing
in a Deep Learning World**

PROCEEDINGS

Edited by Galia Angelova, Ruslan Mitkov, Ivelina Nikolova, Irina Temnikova

Varna, Bulgaria
2–4 September, 2019

**INTERNATIONAL CONFERENCE
RECENT ADVANCES IN
NATURAL LANGUAGE PROCESSING 2019**

Natural Language Processing
in a Deep Learning World

PROCEEDINGS

Varna, Bulgaria
2–8 September 2019

Print ISBN 978-954-452-055-7
Online ISBN 978-954-452-056-4
Series Print ISSN 1313-8502
Series Online ISSN 2603-2813

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

Preface

Welcome to the 12th International Conference on “Recent Advances in Natural Language Processing” (RANLP 2019) in Varna, Bulgaria, 2-4 September 2019. The main objective of the conference is to give researchers the opportunity to present new results in Natural Language Processing (NLP) based on modern theories and methodologies.

The Conference is preceded by the First Summer school on Deep Learning in NLP (29-30 August 2019) and two days of tutorials (31 August – 1 September 2019).

The Summer School lectures are given by Kyunghyun Cho (New York University), Marek Rei (University of Cambridge), Tim Rocktäschel (University College London) and Hinrich Schütze (Ludwig Maximilian University, Munich). Training in practical sessions is provided by Heike Adel (Stuttgart University), Alexander Popov (Institute of Information and Communication Technologies, Bulgarian Academy of Sciences), Omid Rohanian and Shiva Taslimipour (University of Wolverhampton).

Tutorials are given by the following lecturers: Antonio Miceli Barone (University of Edinburgh) and Sheila Castilho (Dublin City University), Valia Kordoni (Humboldt University, Berlin), Preslav Nakov (Qatar Computing Research Institute, HBKU), Vlad Niculae and Tsvetomila Mihaylova (Institute of Telecommunications, Lisbon).

The conference keynote speakers are:

- Kyunghyun Cho (New York University),
- Ken Church (Baidu),
- Preslav Nakov (Qatar Computing Research Institute, HBKU),
- Sebastian Padó (Stuttgart University),
- Hinrich Schütze (Ludwig Maximilian University, Munich).

This year 18 regular papers, 37 short papers, 95 posters, and 7 demos have been accepted for presentation at the conference. The selection rate of accepted papers is: regular papers 8,7%, short papers 26,7%, posters and demo papers – 72%.

The proceedings cover a wide variety of NLP topics, including but not limited to: deep learning; machine translation; opinion mining and sentiment analysis; semantics and discourse; named entity recognition; coreference resolution; corpus annotation; parsing and morphology; text summarisation and simplification; event extraction; fact checking and rumour analysis; NLP for healthcare; and NLP for social media.

In 2019 RANLP hosts four post-conference workshops on influential NLP topics: the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019), the

12th Workshop on Building and Using Comparable Corpora (BUCC), the Multiling 2019 Workshop: Summarization Across Languages, Genres and Sources as well as an Workshop on Language Technology for Digital Historical Archives with a Special Focus on Central-, (South-)Eastern Europe, Middle East and North Africa. The International Conference Biographical Data in a Digital World 2019 is another event held on 5-6 September 2019 in parallel with the RANLP post-conference Workshops.

We would like to thank all members of the Programme Committee and all additional reviewers. Together they have ensured that the best papers were included in the Proceedings and have provided invaluable comments for the authors.

Finally, special thanks go to the University of Wolverhampton, the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences, the Bulgarian National Science Fund, Ontotext and IRIS.AI for their generous support of RANLP.

Welcome to Varna and we hope that you enjoy the conference!

The RANLP 2019 Organisers

The International Conference RANLP–2019 is organised by:

Research Group in Computational Linguistics,
University of Wolverhampton, UK

Linguistic Modelling and Knowledge Processing Department,
Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences, Bulgaria

RANLP–2019 is partially supported by:

National Science Fund, Ministry of Education and Science, Bulgaria

Ontotext AD

IRIS.AI

Programme Committee Chair:

Ruslan Mitkov, University of Wolverhampton, UK

Organising Committee Chair:

Galia Angelova, Bulgarian Academy of Sciences, Bulgaria

Workshop Coordinator:

Kiril Simov, Bulgarian Academy of Sciences, Bulgaria

Tutorial Coordinator:

Preslav Nakov, Qatar Computing Research Institute, HBKU, Qatar

Proceedings Printing:

Nikolai Nikolov, INCOMA Ltd., Shoumen, Bulgaria

Programme Committee Coordinators:

Ivelina Nikolova, Bulgarian Academy of Sciences
Irina Temnikova, Bulgarian Academy of Sciences

Programme Committee:

Ahmed Abdelali (Hamad Bin Khalifa University, Qatar)
Cengiz Acarturk (Middle East Technical University, Turkey)
Guadalupe Aguado-de-Cea (Polytechnic University of Madrid, Spain)
Luis Alfonso Ureña López (University of Jaén, Spain)
Hassina Aliane (Research Center on Scientific and Technical Information, Algeria)
Pascal Amsili (University of Paris Diderot, France)
Galia Angelova (Bulgarian Academy of Sciences, Bulgaria)
Riza Batista-Navarro (University of Manchester, United Kingdom)
Kalina Bontcheva (University of Sheffield, United Kingdom)
Svetla Boytcheva (Bulgarian Academy of Sciences, Bulgaria)
António Branco (University of Lisbon, Portugal)
Chris Brew (Digital Operatives)
Nicoletta Calzolari (Italian National Research Council, Italy)
Sheila Castilho (Dublin City University, Ireland)
Key-Sun Choi (Korea Advanced Institute of Science and Technology, South Korea)
Kenneth Church (Baidu, United States of America)
Kevin Cohen (University of Colorado School of Medicine, United States of America)
Gloria Corpas Pastor (University of Málaga, Spain)
Dan Cristea (University of Iași, Romania)
Antonio Ferrández Rodríguez (University of Alicante, Spain)
Fumiyo Fukumoto (University of Yamanashi, Japan)
Prószéky Gábor (Pázmány Péter Catholic University & Bionics, Hungary)
Alexander Gelbukh (National Polytechnic Institute, Mexico)
Yota Georgakopoulou (Athena Consultancy, Greece)
Ralph Grishman (New York University, United States of America)
Veronique Hoste (Ghent University, Belgium)
Diana Inkpen (University of Ottawa, Canada)
Hitoshi Isahara (Toyohashi University of Technology, Japan)
Miloš Jakubíček (Lexical Computing Ltd)
Alma Kharrat (Microsoft)
Udo Kruschwitz (University of Essex, United Kingdom)
Sandra Kübler (Indiana University, United States of America)
Katia Lida Kermanidis (Ionian University, Greece)
Natalia Loukachevitch (Lomonosov Moscow State University, Russia)
Eid Mohamed (Doha Institute for Graduate Studies, Qatar)
Emad Mohamed (University of Wolverhampton, United Kingdom)
Johanna Monti (University of Naples L'Orientale, Italy)
Andrés Montoyo (University of Alicante, Spain)
Alessandro Moschitti (Amazon)
Rafael Muñoz Guillena (University of Alicante, Spain)
Preslav Nakov (Qatar Computing Research Institute, Qatar)
Roberto Navigli (Sapienza University of Rome, Italy)
Raheel Nawaz (Manchester Metropolitan University, United Kingdom)
Mark-Jan Nederhof (University of St Andrews, United Kingdom)
Ivelina Nikolova (Bulgarian Academy of Sciences, Bulgaria)
Kemal Oflazer (Carnegie Mellon University, Qatar)
Maciej Ogrodniczuk (Polish Academy of Sciences, Poland)

Constantin Orasan (University of Wolverhampton, United Kingdom)
Petya Osenova (Sofia University and Bulgarian Academy of Sciences, Bulgaria)
Sebastian Padó (Stuttgart University, Germany)
Noa P. Cruz Diaz (Artificial Intelligence Excellence Center, Bankia, Spain)
Liviu P. Dinu (University of Bucharest, Romania)
Pavel Pecina (Charles University, Czech Republic)
Stelios Piperidis (Athena Research Center, Greece)
Massimo Poesio (University of Essex, United Kingdom)
Horacio Rodríguez (Polytechnic University of Catalonia, Spain)
Paolo Rosso (Polytechnic University of Valencia, Spain)
Vasile Rus (The University of Memphis, United States of America)
Frédérique Segond (Viseo)
Kiril Simov (Bulgarian Academy of Sciences, Bulgaria)
Vilemini Sosoni (Ionian University, Greece)
Keh-Yih Su (Institute of Information Science, Academia Sinica, Taiwan)
Stan Szpakowicz (University of Ottawa, Canada)
Hristo Tanev (European Commission, Belgium)
Shiva Taslimipoor (University of Wolverhampton, United Kingdom)
Irina Temnikova (Sofia University, Bulgaria)
Dan Tufiş (Romanian Academy of Sciences, Romania)
Aline Villavicencio (University of Essex, United Kingdom
and Federal University of Rio Grande do Sul, Brazil)
Yorick Wilks (Florida Institute for Human and Machine Cognition,
United States of America)
Mai Zaki (American University of Sharjah, United Arab Emirates)
Marcos Zampieri (University of Wolverhampton, United Kingdom)
Michael Zock (University of Aix-Marseille, France)

Reviewers:

Ahmed AbuRa'ed (University Pompeu Fabra, Spain)
Mattia A. Di Gangi (University of Trento, Italy)
Itziar Aldabe (University of País Vasco, Spain)
Ahmed Ali (Hamad Bin Khalifa University, Qatar)
Ahmed Amine Aliane (Research Center on Scientific and Technical Information,
Algeria)
Le An Ha (University of Wolverhampton, United Kingdom)
Atefeh (Anna) Farzindar (University of Southern California, United States of America)
João António Rodrigues (University of Lisboa, Portugal)
Pepa Atanasova (University of Copenhagen, Denmark)
Mohammed Attia (George Washington University, United States of America)
Parnia Bahar (Aachen University, Germany)
Belahcene Bahloul (University of Khemis Miliana, Algeria)
Eduard Barbu (University of Tartu, Estonia)
Alberto Barrón-Cedeño (University of Bologna, Italy)
Leonor Becerra (Jean Monnet University, France)
Andrea Bellandi (National Research Council, Italy)
Fernando Benites (ZHAW School of Engineering, Switzerland)
Victoria Bobicev (Technical University of Moldova, Moldova)

Antonina Bondarenko (Lipetsk State Technical University, Russia)
Aurélien Bossard (University Paris 8, France)
Aljoscha Burchardt (German Research Centre for Artificial Intelligence, Germany)
Lindsay Bywood (University of Westminster, United Kingdom)
Ruket Cakici (Middle East Technical University, Turkey)
Iacer Calixto (University of Amsterdam, Netherlands and New York University, United States of America)
Pablo Calleja (Polytechnic University of Madrid, Spain)
Erik Cambria (Nanyang Technological University, Singapore)
Kai Cao (New York University, United States of America)
Thiago Castro Ferreira (Tilburg University, Netherlands)
Yue Chen (Queen Mary University of London, United Kingdom)
Mihaela Colhon (University of Craiova, Romania)
Daniel Dakota (Indiana University, United States of America)
Kareem Darwish (Hamad Bin Khalifa University, Qatar)
Orphee De Clercq (Ghent University, Belgium)
Kevin Deturck (Viseo)
Asma Djaidri (University of Science and Technology Houari Boumediene, Algeria)
Mazen Elagamy (Staffordshire University, United Kingdom)
Can Erten (University of York, United Kingdom)
Luis Espinosa Anke (Cardiff University, United Kingdom)
Kilian Evang (University of Düsseldorf, Germany)
Richard Evans (University of Wolverhampton, United Kingdom)
Stefan Evert (Friedrich–Alexander University, Germany)
Anna Feherova (University of Wolverhampton, United Kingdom)
Mariano Felice (University of Cambridge, United Kingdom)
Corina Forascu (The Alexandru Ioan Cuza University, Romania)
Vasiliki Foufi (University of Geneva, Switzerland)
Thomas Francois (Université catholique de Louvain, Belgium)
Adam Funk (University of Sheffield, United Kingdom)
Björn Gambäck (Norwegian University of Science and Technology, Norway)
Aina Garí Soler (The Computer Science Laboratory for Mechanics and Engineering Sciences, France)
Federico Gaspari (Dublin City University, Ireland)
José G. C. de Souza (eBay)
Goran Glavaš (University of Mannheim, Germany)
Darina Gold (University of Duisburg-Essen, Germany)
Reshmi Gopalakrishna Pillai (University of Wolverhampton, United Kingdom)
Rohit Gupta (University of Wolverhampton, United Kingdom)
Amir Hazem (Nantes University, France)
Tomáš Hercig (University of West Bohemia, Czech Republic)
Yasser Hifny (University of Helwan, Egypt)
Diliara Iakubova (Kazan Federal University, Russia)
Adrian Iftene (The Alexandru Ioan Cuza University, Romania)
Camelia Ignat (European Commission, Belgium)
Dmitry Ilvovsky (National Research University Higher School of Economics, Russia)
Miloš Jakubíček (Masaryk University, Czech Republic)
Arkadiusz Janz (Wroclaw University of Science and Technology, Poland)

Héctor Jiménez-Salazar (The Metropolitan Autonomous University, Mexico)
Olga Kanishcheva (National Technical University, Ukraine)
Georgi Karadzhov (Sofia University, Bulgaria)
David Kauchak (Pomona College, United States of America)
Yasen Kiproff (Sofia University, Bulgaria)
Jan Kocoń (Wroclaw University of Science and Technology, Poland)
Sarah Kohail (Hamburg University, Germany)
Yannis Korkontzelos (Edge Hill University, United Kingdom)
Venelin Kovatchev (University of Barcelona, Spain)
Peter Krejzl (University of West Bohemia, Czech Republic)
Sudip Kumar Naskar (Jadavpur University, India)
Maria Kunilovskaya (University of Tyumen, Russia)
Andrey Kutuzov (University of Oslo, Norway)
Sobha Lalitha Devi (Anna University, India)
Gabriella Lapesa (Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, Germany)
Todor Lazarov (Bulgarian Academy of Sciences, Bulgaria)
Els Lefever (Ghent University, Belgium)
Ladislav Lenc (University of West Bohemia, Czech Republic)
Elena Lloret (University of Alicante)
Pintu Lohar (Dublin City University, Ireland)
Epida Loupaki (Aristotle University of Thessaloniki, Greece)
Lieve Macken (Ghent University, Belgium)
Mireille Makary (University of Wolverhampton, United Kingdom)
Michał Marcińczuk (Wroclaw University of Technology, Poland)
Angelo Mario Del Grosso (National Research Council of Italy, Italy)
Federico Martelli (Babelscape, Italy)
Patricia Martín Chozas (Polytechnic University of Madrid, Spain)
Eugenio Martínez-Cámara (University of Granada, Spain)
Irina Matveeva (NexLP, United States of America)
Flor Miriam Plaza del Arco (University of Jaén, Spain)
Arturo Montejo-Ráez (University of Jaén, Spain)
Paloma Moreda Pozo (University of Alicante, Spain)
Diego Moussallem (University of Paderborn, Germany)
Sara Moze (University of Wolverhampton, United Kingdom)
Nona Naderi (University of Toronto, Canada)
Marcin Oleksy (Wrocław University of Science and Technology, Poland)
Antoni Oliver (The Open University of Catalonia, Spain)
Mihaela Onofrei (University of Iasi, Romania)
Arzucan Özgür (Bogazici University, Turkey)
Santanu Pal (Saarland University, Germany)
Alexander Panchenko (University of Hamburg, Germany)
Sean Papay (University of Stuttgart, Germany)
Ljudmila Petković (University of Belgrade, Serbia)
Maciej Piasecki (Wroclaw University of Science and Technology, Poland)
Paul Piwek (The Open University, United Kingdom)
Alistair Plum (University of Wolverhampton, United Kingdom)
Alberto Poncelas (Dublin City University, Ireland)

Alexander Popov (Bulgarian Academy of Sciences, Bulgaria)
Maja Popović (Dublin City University, Ireland)
Dan Povey (Johns Hopkins University, United States of America)
Ondřej Pražák (University of West Bohemia, Czech Republic)
Prokopis Prokopidis (Research and Innovation Center in Information, Greece)
Tharindu Ranasinghe (University of Wolverhampton, United Kingdom)
Natalia Resende (Dublin City University, Ireland)
Pattabhi RK Rao (Anna University, India)
Omid Rohanian (University of Wolverhampton, United Kingdom)
Josef Ruppenhofer (Institute for the German Language, Germany)
Pavel Rychlý (Masaryk University, Czech Republic)
Magdaléna Rysová (Charles University, Czech Republic)
Branislava Šandrih (Belgrade University, Serbia)
Estela Saquete (University of Alicante, Spain)
Leah Schaede (Indiana University, United States)
Ineke Schuurman (University of Leuven, Belgium)
Olga Seminck (Paris Diderot University, France)
Nasredine Semmar (Laboratory for Integration of Systems and Technology, France)
Matthew Shardlow (Manchester Metropolitan University, United Kingdom)
Artem Shelmanov (Russian Academy of Sciences, Russia)
Dimitar Shterionov (Dublin City University, Ireland)
Jennifer Sikos (University of Stuttgart, Germany)
João Silva (University of Lisboa, Portugal)
Vasiliki Simaki (Lancaster University, United Kingdom)
Sunayana Sitaram (Microsoft Research, India)
Mihailo Skoric (Researcher, Serbia)
Felix Stahlberg (University of Cambridge, Department of Engineering, United Kingdom)
Kenneth Steimel (Indiana University, United States)
Sebastian Stüker (Karlsruhe Institute of Technology, Germany)
Yoshimi Suzuki (Shizuoka University, Japan)
Liling Tan (Nanyang Technological University, Singapore)
Segun Taofeek Aroyehun (National Polytechnic Institute, Mexico)
Laura Tološi (Self employed data scientist)
Elena Tutubalina (Kazan Federal University, Russia)
Eleni Tziafa (National and Kapodistrian University of Athens, Greece)
Antonio Valerio Miceli Barone (University of Edinburgh, United Kingdom)
Mihaela Vela (Saarland University, Germany)
Cristina Vertan (University of Hamburg, Germany)
Manuel Vilares Ferro (University of Vigo, Spain)
Veronika Vincze (University of Szeged, Hungary)
Pidong Wang (National University of Singapore, Singapore)
Michael Wiegand (Heidelberg University, Germany)
Victoria Yaneva (University of Wolverhampton, United Kingdom)
Kristina Yordanova (University of Rostock, Germany)
Juntao Yu (Queen Mary University of London, United Kingdom)
Wajdi Zaghouni (Hamad Bin Khalifa University, Qatar)
Kalliopi Zervanou (Eindhoven University of Technology, Netherlands)
Inès Zribi (University of Sfax, Tunisia)

Table of Contents

<i>Table Structure Recognition Based on Cell Relationship, a Bottom-Up Approach</i> Darshan Adiga, Shabir Ahmad Bhat, Muzaffar Bashir Shah and Viveka Vyeth	1
<i>Identification of Good and Bad News on Twitter</i> Piush Aggarwal and Ahmet Aker	9
<i>Bilingual Low-Resource Neural Machine Translation with Round-Tripping: The Case of Persian-Spanish</i> Benyamin Ahmadnia and Bonnie Dorr	18
<i>Enhancing Phrase-Based Statistical Machine Translation by Learning Phrase Representations Using Long Short-Term Memory Network</i> Benyamin Ahmadnia and Bonnie Dorr	25
<i>Automatic Propbank Generation for Turkish</i> Koray Ak and Olcay Taner Yıldız	33
<i>Multilingual Sentence-Level Bias Detection in Wikipedia</i> Desislava Aleksandrova, François Lareau and Pierre André Ménard	42
<i>Supervised Morphological Segmentation Using Rich Annotated Lexicon</i> Ebrahim Ansari, Zdeněk Žabokrtský, Mohammad Mahmoudi, Hamid Haghdoost and Jonáš Vidra	52
<i>Combining Lexical Substitutes in Neural Word Sense Induction</i> Nikolay Arefyev, Boris Sheludko and Alexander Panchenko	62
<i>Detecting Clitics Related Orthographic Errors in Turkish</i> Uğurcan Arıkan, Onur Güngör and Suzan Uskudarlı	71
<i>Benchmark Dataset for Propaganda Detection in Czech Newspaper Texts</i> Vít Baisa, Ondřej Herman and Ales Horak	77
<i>Diachronic Analysis of Entities by Exploiting Wikipedia Page Revisions</i> Pierpaolo Basile, Annalina Caputo, Seamus Lawless and Giovanni Semeraro	84
<i>Using a Lexical Semantic Network for the Ontology Building</i> Nadia Bebashina-Clairet, Sylvie Despres and Mathieu Lafourcade	92
<i>Naive Regularizers for Low-Resource Neural Machine Translation</i> Meriem Beloucif, Ana Valeria Gonzalez, Marcel Bollmann and Anders Søgaard	102
<i>Exploring Graph-Algebraic CCG Combinators for Syntactic-Semantic AMR Parsing</i> Sebastian Beschke	112
<i>Quasi Bidirectional Encoder Representations from Transformers for Word Sense Disambiguation</i> Michele Bevilacqua and Roberto Navigli	122
<i>Evaluating the Consistency of Word Embeddings from Small Data</i> Jelke Bloem, Antske Fokkens and Aurélie Herbelot	132
<i>Cross-Domain Training for Goal-Oriented Conversational Agents</i> Alexandra Maria Bodîrlău, Stefania Budulan and Traian Rebedea	142

<i>Learning Sentence Embeddings for Coherence Modelling and Beyond</i> Tanner Bohn, Yining Hu, Jinhang Zhang and Charles Ling	151
<i>Risk Factors Extraction from Clinical Texts Based on Linked Open Data</i> Svetla Boytcheva, Galia Angelova and Zhivko Angelov	161
<i>Parallel Sentence Retrieval From Comparable Corpora for Biomedical Text Simplification</i> Rémi Cardon and Natalia Grabar	168
<i>Classifying Author Intention for Writer Feedback in Related Work</i> Arlene Casey, Bonnie Webber and Dorota Glowacka	178
<i>Sparse Victory – A Large Scale Systematic Comparison of Count-Based and Prediction-Based Vectorizers for Text Classification</i> Rupak Chakraborty, Ashima Elhence and Kapil Arora	188
<i>A Fine-Grained Annotated Multi-Dialectal Arabic Corpus</i> Anis Charfi, Wajdi Zaghouani, Syed Hassan Mehdi and Esraa Mohamed	198
<i>Personality-Dependent Neural Text Summarization</i> Pablo Costa and Ivandré Paraboni	205
<i>Self-Adaptation for Unsupervised Domain Adaptation</i> Xia Cui and Danushka Bollegala	213
<i>Speculation and Negation Detection in French Biomedical Corpora</i> Clément Dalloux, Vincent Claveau and Natalia Grabar	223
<i>Porting Multilingual Morphological Resources to OntoLex-Lemon</i> Thierry Declerck and Stefania Racioppa	233
<i>Dependency-Based Self-Attention for Transformer NMT</i> Hiroyuki Deguchi, Akihiro Tamura and Takashi Ninomiya	239
<i>Detecting Toxicity in News Articles: Application to Bulgarian</i> Yoan Dinkov, Ivan Koychev and Preslav Nakov	247
<i>De-Identification of Emails: Pseudonymizing Privacy-Sensitive Data in a German Email Corpus</i> Elisabeth Eder, Ulrike Krieg-Holz and Udo Hahn	259
<i>Lexical Quantile-Based Text Complexity Measure</i> Maksim Ereemeev and Konstantin Vorontsov	270
<i>Demo Application for LETO: Learning Engine Through Ontologies</i> Suilan Estevez-Velarde, Andrés Montoyo, Yudivian Almeida-Cruz, Yoan Gutiérrez, Alejandro Piad-Morffis and Rafael Muñoz	276
<i>Sentence Simplification for Semantic Role Labelling and Information Extraction</i> Richard Evans and Constantin Orăsan	285
<i>OlloBot - Towards A Text-Based Arabic Health Conversational Agent: Evaluation and Results</i> Ahmed Fadhil and Ahmed AbuRa'ed	295
<i>Developing the Old Tibetan Treebank</i> Christian Faggionato and Marieke Meelen	304

<i>Summarizing Legal Rulings: Comparative Experiments</i> Diego Feijo and Viviane Moreira	313
<i>Entropy as a Proxy for Gap Complexity in Open Cloze Tests</i> Mariano Felice and Paula Buttery	323
<i>Song Lyrics Summarization Inspired by Audio Thumbnailing</i> Michael Fell, Elena Cabrio, Fabien Gandon and Alain Giboin	328
<i>Comparing Automated Methods to Detect Explicit Content in Song Lyrics</i> Michael Fell, Elena Cabrio, Michele Corazza and Fabien Gandon	338
<i>Linguistic Classification: Dealing Jointly with Irrelevance and Inconsistency</i> Laura Franzoi, Andrea Sgarro, Anca Dinu and Liviu P. Dinu	345
<i>Corpus Lexicography in a Wider Context</i> Chen Gafni	353
<i>A Universal System for Automatic Text-to-Phonetics Conversion</i> Chen Gafni	360
<i>Two Discourse Tree - Based Approaches to Indexing Answers</i> Boris Galitsky and Dmitry Ilvovsky	367
<i>Discourse-Based Approach to Involvement of Background Knowledge for Question Answering</i> Boris Galitsky and Dmitry Ilvovsky	373
<i>On a Chatbot Providing Virtual Dialogues</i> Boris Galitsky, Dmitry Ilvovsky and Elizaveta Goncharova	382
<i>Assessing Socioeconomic Status of Twitter Users: A Survey</i> Dhouha Ghazouani, Luigi Lancieri, Habib Ounelli and Chaker Jebari	388
<i>Divide and Extract – Disentangling Clause Splitting and Proposition Extraction</i> Darina Gold and Torsten Zesch	399
<i>Sparse Coding in Authorship Attribution for Polish Tweets</i> Piotr Grzybowski, Ewa Juralewicz and Maciej Piasecki	409
<i>Automatic Question Answering for Medical MCQs: Can It Go Further than Information Retrieval?</i> Le An Ha and Victoria Yaneva	418
<i>Self-Knowledge Distillation in Natural Language Processing</i> Sangchul Hahn and Heeyoul Choi	424
<i>From the Paft to the Fiiture: A Fully Automatic NMT and Word Embeddings Method for OCR Post-Correction</i> Mika Hämäläinen and Simon Hengchen	432
<i>Investigating Terminology Translation in Statistical and Neural Machine Translation: A Case Study on English-to-Hindi and Hindi-to-English</i> Rejwanul Haque, Md Hasanuzzaman and Andy Way	438
<i>Beyond English-Only Reading Comprehension: Experiments in Zero-Shot Multilingual Transfer for Bulgarian</i> Momchil Hardalov, Ivan Koychev and Preslav Nakov	448

<i>Tweaks and Tricks for Word Embedding Disruptions</i>	
Amir Hazem and Nicolas Hernandez	461
<i>Meta-Embedding Sentence Representation for Textual Similarity</i>	
Amir Hazem and Nicolas Hernandez	466
<i>Emoji Powered Capsule Network to Detect Type and Target of Offensive Posts in Social Media</i>	
Hansi Hettiarachchi and Tharindu Ranasinghe	475
<i>EoANN: Lexical Semantic Relation Classification Using an Ensemble of Artificial Neural Networks</i>	
Rayehe Hosseini Pour and Mehrnoush Shamsfard	482
<i>Opinions Summarization: Aspect Similarity Recognition Relaxes the Constraint of Predefined Aspects</i>	
Nguyen Huy Tien, Le Tung Thanh and Nguyen Minh Le	488
<i>Discourse-Aware Hierarchical Attention Network for Extractive Single-Document Summarization</i>	
Tatsuya Ishigaki, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura	498
<i>Semi-Supervised Induction of POS-Tag Lexicons with Tree Models</i>	
Maciej Janicki	508
<i>Word Sense Disambiguation Based on Constrained Random Walks in Linked Semantic Networks</i>	
Arkadiusz Janz and Maciej Piasecki	517
<i>Classification of Micro-Texts Using Sub-Word Embeddings</i>	
Mihir Joshi and Nur Zincir-Heywood	527
<i>Using Syntax to Resolve NPE in English</i>	
Payal Khullar, Allen Antony and Manish Shrivastava	535
<i>Is Similarity Visually Grounded? Computational Model of Similarity for the Estonian Language</i>	
Claudia Kittask and Eduard Barbu	542
<i>Language-Agnostic Twitter-Bot Detection</i>	
Jürgen Knauth	551
<i>Multi-Level Analysis and Recognition of the Text Sentiment on the Example of Consumer Opinions</i>	
Jan Kocoń, Monika Zaśko-Zielińska and Piotr Miłkowski	560
<i>A Qualitative Evaluation Framework for Paraphrase Identification</i>	
Venelin Kovatchev, M. Antònia Martí, Maria Salamo and Javier Beltran	569
<i>Study on Unsupervised Statistical Machine Translation for Backtranslation</i>	
Anush Kumar, Nihal V. Nayak, Aditya Chandra and Mydhili K. Nair	579
<i>Towards Functionally Similar Corpus Resources for Translation</i>	
Maria Kunilovskaya and Serge Sharoff	584
<i>Question Similarity in Community Question Answering: A Systematic Exploration of Preprocessing Methods and Models</i>	
Florian Kunneman, Thiago Castro Ferreira, Emiel Kraemer and Antal van den Bosch	594
<i>A Classification-Based Approach to Cognate Detection Combining Orthographic and Semantic Similarity Information</i>	
Sofie Labat and Els Lefever	603

<i>Resolving Pronouns for a Resource-Poor Language, Malayalam Using Resource-Rich Language, Tamil.</i> Sobha Lalitha Devi	612
<i>Semantic Role Labeling with Pretrained Language Models for Known and Unknown Predicates</i> Daniil Larionov, Artem Shelmanov, Elena Chistova and Ivan Smirnov	620
<i>A Structural Approach to Enhancing WordNet with Conceptual Frame Semantics</i> Svetlozara Leseva and Ivelina Stoyanova.....	630
<i>Compositional Hyponymy with Positive Operators</i> Martha Lewis	639
<i>The Impact of Semantic Linguistic Features in Relation Extraction: A Logical Relational Learning Approach</i> Rinaldo Lima, Bernard Espinasse and Frederico Freitas	649
<i>Detecting Anorexia in Spanish Tweets</i> Pilar López Úbeda, Flor Miriam Plaza del Arco, Manuel Carlos Díaz Galiano, L. Alfonso Urena Lopez and Maite Martin	656
<i>A Type-Theoretical Reduction of Morphological, Syntactic and Semantic Compositionality to a Single Level of Description</i> Erkki Luuk	665
<i>v-trel: Vocabulary Trainer for Tracing Word Relations - An Implicit Crowdsourcing Approach</i> Verena Lyding, Christos Rodosthenous, Federico Sangati, Umair ul Hassan, Lionel Nicolas, Alexander König, Jolita Horbacauskiene and Anisia Katinskaia.....	675
<i>Jointly Learning Author and Annotated Character N-gram Embeddings: A Case Study in Literary Text</i> Suraj Maharjan, Deepthi Mave, Prasha Shrestha, Manuel Montes, Fabio A. González and Thamar Solorio	685
<i>Generating Challenge Datasets for Task-Oriented Conversational Agents through Self-Play</i> Sourabh Majumdar, Serra Sinem Tekiroglu and Marco Guerini.....	694
<i>Sentiment Polarity Detection in Azerbaijani Social News Articles</i> Sevda Mammadli, Shamsaddin Huseynov, Huseyn Alkaramov, Ulviyya Jafarli, Umid Suleymanov and Samir Rustamov.....	704
<i>Inforex — a Collaborative System for Text Corpora Annotation and Analysis Goes Open</i> Michał Marcińczuk and Marcin Oleksy	712
<i>Semantic Language Model for Tunisian Dialect</i> Abir Masmoudi, Rim Laatar, Mariem Ellouze and Lamia Hadrich Belguith.....	721
<i>Automatic Diacritization of Tunisian Dialect Text Using Recurrent Neural Network</i> Abir Masmoudi, Mariem Ellouze and Lamia Hadrich Belguith	731
<i>Comparing MT Approaches for Text Normalization</i> Claudia Matos Veliz, Orphee De Clercq and Veronique Hoste	741
<i>Sentiment and Emotion Based Representations for Fake Reviews Detection</i> Alimuddin Melleng, Anna Jurek-Loughrey and Deepak P.....	751
<i>Turning Silver into Gold: Error-Focused Corpus Reannotation with Active Learning</i> Pierre André Ménard and Antoine Mougeot	759

<i>NLP Community Perspectives on Replicability</i>	
Margot Mieskes, Karën Fort, Aurélie Névéol, Cyril Grouin and Kevin Cohen	769
<i>Unsupervised Data Augmentation for Less-Resourced Languages with no Standardized Spelling</i>	
Alice Millour and Karën Fort	777
<i>Neural Feature Extraction for Contextual Emotion Detection</i>	
Elham Mohammadi, Hessam Amini and Leila Kosseim	786
<i>Empirical Study of Diachronic Word Embeddings for Scarce Data</i>	
Syrielle Montariol and Alexandre Allauzen	796
<i>A Fast and Accurate Partially Deterministic Morphological Analysis</i>	
Hajime Morita and Tomoya Iwakura	805
<i>incom.py - A Toolbox for Calculating Linguistic Distances and Asymmetries between Related Languages</i>	
Marius Mosbach, Irina Stenger, Tania Avgustinova and Dietrich Klakow	811
<i>A Holistic Natural Language Generation Framework for the Semantic Web</i>	
Axel-Cyrille Ngonga Ngomo, Diego Moussallem and Lorenz Bühmann	820
<i>Building a Comprehensive Romanian Knowledge Base for Drug Administration</i>	
Bogdan Nicula, Mihai Dascalu, Maria-Dorinela Sîrbu, Ștefan Trăușan-Matu and Alexandru Nuță	830
<i>Summary Refinement through Denoising</i>	
Nikola Nikolov, Alessandro Calmanovici and Richard Hahnloser	838
<i>Large-Scale Hierarchical Alignment for Data-Driven Text Rewriting</i>	
Nikola Nikolov and Richard Hahnloser	845
<i>Dependency-Based Relative Positional Encoding for Transformer NMT</i>	
Yutaro Omote, Akihiro Tamura and Takashi Ninomiya	855
<i>From Image to Text in Sentiment Analysis via Regression and Deep Learning</i>	
Daniela Onita, Liviu P. Dinu and Adriana Birlutiu	863
<i>Building a Morphological Analyser for Laz</i>	
Esra Önal and Francis Tyers	870
<i>Term Based Semantic Clusters for Very Short Text Classification</i>	
Jasper Paalman, Shantanu Mullick, Kalliopi Zervanou, Yingqian Zhang	879
<i>Quotation Detection and Classification with a Corpus-Agnostic Model</i>	
Sean Papay and Sebastian Padó	889
<i>Validation of Facts Against Textual Sources</i>	
Vamsi Krishna Pendyala, Simran Sinha, Satya Prakash, Shriya Reddy and Anupam Jamatia	896
<i>A Neural Network Component for Knowledge-Based Semantic Representations of Text</i>	
Alejandro Piad-Morffis, Rafael Muñoz, Yudivian Almeida-Cruz, Yoan Gutiérrez, Suilan Estevez-Velarde and Andrés Montoyo	905
<i>Toponym Detection in the Bio-Medical Domain: A Hybrid Approach with Deep Learning</i>	
Alistair Plum, Tharindu Ranasinghe and Constantin Orăsan	913

<i>Combining SMT and NMT Back-Translated Data for Efficient NMT</i> Alberto Poncelas, Maja Popović, Dimitar Shterionov, Gideon Maillette de Buy Wenniger and Andy Way	923
<i>Unsupervised Dialogue Intent Detection via Hierarchical Topic Model</i> Artem Popov, Victor Bulatov, Darya Polyudova and Eugenia Veselova	933
<i>Graph Embeddings for Frame Identification</i> Alexander Popov and Jennifer Sikos	940
<i>Know Your Graph. State-of-the-Art Knowledge-Based WSD</i> Alexander Popov, Kiril Simov and Petya Osenova	950
<i>Are Ambiguous Conjunctions Problematic for Machine Translation?</i> Maja Popović and Sheila Castilho	960
<i>ULSAna: Universal Language Semantic Analyzer</i> Ondřej Pražák and Miloslav Konopík	968
<i>Machine Learning Approach to Fact-Checking in West Slavic Languages</i> Pavel Přibáň, Tomáš Hercig and Josef Steinberger	974
<i>NE-Table: A Neural Key-Value Table for Named Entities</i> Janarthanan Rajendran, Jatin Ganhotra, Xiaoxiao Guo, Mo Yu, Satinder Singh and Lazaros Polymenakos	981
<i>Enhancing Unsupervised Sentence Similarity Methods with Deep Contextualised Word Representations</i> Tharindu Ranasinghe, Constantin Orăsan and Ruslan Mitkov	995
<i>Semantic Textual Similarity with Siamese Neural Networks</i> Tharindu Ranasinghe, Constantin Orăsan and Ruslan Mitkov	1005
<i>Analysing the Impact of Supervised Machine Learning on Automatic Term Extraction: HAMLET vs TermoStat</i> Ayla Rigouts Terryn, Patrick Drouin, Veronique Hoste and Els Lefever	1013
<i>Distant Supervision for Sentiment Attitude Extraction</i> Nicolay Rusnachenko, Natalia Loukachevitch and Elena Tutubalina	1023
<i>Self-Attentional Models Application in Task-Oriented Dialogue Generation Systems</i> Mansour Saffar Mehrjardi, Amine Trabelsi and Osmar R. Zaiane	1032
<i>Whom to Learn From? Graph- vs. Text-Based Word Embeddings</i> Małgorzata Salawa, António Branco, Ruben Branco, João António Rodrigues and Chakaveh Saedi	1042
<i>Persistence Pays Off: Paying Attention to What the LSTM Gating Mechanism Persists</i> Giancarlo Salton and John Kelleher	1053
<i>Development and Evaluation of Three Named Entity Recognition Systems for Serbian - The Case of Personal Names</i> Branislava Šandrih, Cvetana Krstev and Ranka Stankovic	1061
<i>Moral Stance Recognition and Polarity Classification from Twitter and Elicited Text</i> Wesley Santos and Ivandré Paraboni	1070

<i>The "Jump and Stay" Method to Discover Proper Verb Centered Constructions in Corpus Lattices</i> Bálint Sass	1077
<i>Offence in Dialogues: A Corpus-Based Study</i> Johannes Schäfer and Ben Burtenshaw	1086
<i>EmoTag – Towards an Emotion-Based Analysis of Emojis</i> Abu Awal Md Shoeb, Shahab Raji and Gerard de Melo	1095
<i>A Morpho-Syntactically Informed LSTM-CRF Model for Named Entity Recognition</i> Lilia Simeonova, Kiril Simov, Petya Osenova and Preslav Nakov	1105
<i>Named Entity Recognition in Information Security Domain for Russian</i> Anastasiia Sirotina and Natalia Loukachevitch	1115
<i>Cross-Family Similarity Learning for Cognate Identification in Low-Resource Languages</i> Elieel Soisalon-Soininen and Mark Granroth-Wilding	1122
<i>Automatic Detection of Translation Direction</i> Iliia Sominsky and Shuly Wintner	1132
<i>Automated Text Simplification as a Preprocessing Step for Machine Translation into an Under-Resourced Language</i> Sanja Štajner and Maja Popović	1142
<i>Investigating Multilingual Abusive Language Detection: A Cautionary Tale</i> Kenneth Steimel, Daniel Dakota, Yue Chen and Sandra Kübler	1152
<i>Augmenting a BiLSTM Tagger with a Morphological Lexicon and a Lexical Category Identification Step</i> Steinþór Steingrímsson, Örvar Káráson and Hrafn Loftsson	1162
<i>Comparison of Machine Learning Approaches for Industry Classification Based on Textual Descriptions of Companies</i> Andrey Tagarev, Nikola Tulechki and Svetla Boytcheva	1170
<i>A Quantum-Like Approach to Word Sense Disambiguation</i> Fabio Tamburini	1177
<i>Understanding Neural Machine Translation by Simplification: The Case of Encoder-Free Models</i> Gongbo Tang, Rico Sennrich and Joakim Nivre	1187
<i>Text-Based Joint Prediction of Numeric and Categorical Attributes of Entities in Knowledge Bases</i> V Thejas, Abhijeet Gupta and Sebastian Padó	1195
<i>SenZi: A Sentiment Analysis Lexicon for the Latinised Arabic (Arabizi)</i> Taha Tobaili, Miriam Fernandez, Harith Alani, Sanaa Sharafeddine, Hazem Hajj and Goran Glavaš	1204
<i>Mining the UK Web Archive for Semantic Change Detection</i> Adam Tsakalidis, Marya Bazzi, Mihai Cucuringu, Pierpaolo Basile and Barbara McGillivray	1213
<i>Cross-Lingual Word Embeddings for Morphologically Rich Languages</i> Ahmet Üstün, Gosse Bouma and Gertjan van Noord	1223
<i>It Takes Nine to Smell a Rat: Neural Multi-Task Learning for Check-Worthiness Prediction</i> Slavena Vasileva, Pepa Atanasova, Lluís Màrquez, Alberto Barrón-Cedeño and Preslav Nakov	1230

<i>Deep Learning Contextual Models for Prediction of Sport Events Outcome from Sportsmen Interviews</i> Boris Velichkov, Ivan Koychev and Svetla Boytcheva	1241
<i>Exploiting Frame-Semantics and Frame-Semantic Parsing for Automatic Extraction of Typological Information from Descriptive Grammars of Natural Languages</i> Shafqat Mumtaz Virk, Azam Sheikh Muhammad, Lars Borin, Muhammad Irfan Aslam, Saania Iqbal and Nazia Khurram	1248
<i>Exploiting Open IE for Deriving Multiple Premises Entailment Corpus</i> Martin Vítá and Jakub Klímeš	1258
<i>Towards Adaptive Text Summarization: How Does Compression Rate Affect Summary Readability of L2 Texts?</i> Tatiana Vodolazova and Elena Lloret	1266
<i>The Impact of Rule-Based Text Generation on the Quality of Abstractive Summaries</i> Tatiana Vodolazova and Elena Lloret	1276
<i>ETNLP: A Visual-Aided Systematic Approach to Select Pre-Trained Embeddings for a Downstream Task</i> Son Vu Xuan, Thanh Vu, Son Tran and Lili Jiang	1286
<i>Tagger for Polish Computer Mediated Communication Texts</i> Wiktor Walentynowicz, Maciej Piasecki and Marcin Oleksy	1296
<i>Evaluation of Vector Embedding Models in Clustering of Text Documents</i> Tomasz Walkowiak and Mateusz Gniewkowski	1305
<i>Bigger versus Similar: Selecting a Background Corpus for First Story Detection Based on Distributional Similarity</i> Fei Wang, Robert J. Ross and John D. Kelleher	1313
<i>Predicting Sentiment of Polish Language Short Texts</i> Aleksander Wawer and Julita Sobiczewska	1322
<i>Improving Named Entity Linking Corpora Quality</i> Albert Weichselbraun, Adrian M.P. Brasoveanu, Philipp Kuntschik and Lyndon J.B. Nixon ..	1329
<i>Sequential Graph Dependency Parser</i> Sean Welleck and Kyunghyun Cho	1339
<i>Term-Based Extraction of Medical Information: Pre-Operative Patient Education Use Case</i> Martin Wolf, Volha Petukhova and Dietrich Klakow	1347
<i>A Survey of the Perceived Text Adaptation Needs of Adults with Autism</i> Victoria Yaneva, Constantin Orăsan, Le An Ha and Natalia Ponomareva	1357
<i>An Open, Extendible, and Fast Turkish Morphological Analyzer</i> Olcay Taner Yıldız, Begüm Avar and Gökhan Ercan	1365
<i>Self-Attention Networks for Intent Detection</i> Sevinj Yolchuyeva, Géza Németh and Bálint Gyires-Tóth	1374
<i>Turkish Tweet Classification with Transformer Encoder</i> Atif Emre Yüksel, Yaşar Alim Türkmen, Arzucan Özgür and Berna Altunel	1381

<i>Multilingual Dynamic Topic Model</i>	
Elaine Zosa and Mark Granroth-Wilding	1389
<i>A Wide-Coverage Context-Free Grammar for Icelandic and an Accompanying Parsing System</i>	
Vilhjalmur Þorsteinsson, Hulda Óladóttir and Hrafn Loftsson	1398