

Table of Contents – Part I

General Techniques

Unsupervised Feature Adaptation for Cross-Domain NLP with an Application to Compositionality Grading	1
<i>Lukas Michelbacher, Qi Han, and Hinrich Schütze</i>	
Syntactic Dependency-Based N-grams: More Evidence of Usefulness in Classification	13
<i>Grigori Sidorov, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, and Liliana Chanona-Hernández</i>	

Lexical Resources

A Quick Tour of BabelNet 1.1 (Invited paper)	25
<i>Roberto Navigli</i>	
Automatic Pipeline Construction for Real-Time Annotation	38
<i>Henning Wachsmuth, Mirko Rose, and Gregor Engels</i>	
A Multilingual GRUG Treebank for Underresourced Languages	50
<i>Oleg Kapanadze and Alla Mishchenko</i>	
Creating an Annotated Corpus for Extracting Canonical Citations from Classics-Related Texts by Using Active Annotation	60
<i>Matteo Romanello</i>	
Approaches of Anonymisation of an SMS Corpus	77
<i>Namrata Patel, Pierre Accorsi, Diana Inkpen, Cédric Lopez, and Mathieu Roche</i>	
A Corpus Based Approach for the Automatic Creation of Arabic Broken Plural Dictionaries	89
<i>Samhaa R. El-Beltagy and Ahmed Rafea</i>	
Temporal Classifiers for Predicting the Expansion of Medical Subject Headings	98
<i>George Tsatsaronis, Iraklis Varlamis, Nattiya Kanhabua, and Kjetil Nørvang</i>	
Knowledge Discovery on Incompatibility of Medical Concepts	114
<i>Adam Grycner, Patrick Ernst, Amy Siu, and Gerhard Weikum</i>	
Extraction of Part-Whole Relations from Turkish Corpora	126
<i>Tuğba Yıldız, Savaş Yıldırım, and Banu Diri</i>	

Chinese Terminology Extraction Using EM-Based Transfer Learning Method	139
<i>Yanxia Qin, Dequan Zheng, Tiejun Zhao, and Min Zhang</i>	
Orthographic Transcription for Spoken Tunisian Arabic	153
<i>Inès Zribi, Marwa Graja, Mariem Ellouze Khmekhem, Maher Jaoua, and Lamia Hadrich Belguith</i>	

Morphology and Tokenization

An Improved Stemming Approach Using HMM for a Highly Inflectional Language	164
<i>Navanath Saharia, Kishori M. Konwar, Utpal Sharma, and Jugul K. Kalita</i>	
Semi-automatic Acquisition of Two-Level Morphological Rules for Iban Language	174
<i>Suhaila Saeed, Lay-Ki Soon, Tek Yong Lim, Bali Ranaivo-Malançon, and Enya Kong Tang</i>	
Finite State Morphology for Amazigh Language	189
<i>Fatima Zahra Nejme, Siham Boulaknadel, and Driss Aboutajdine</i>	
New Perspectives in Sinographic Language Processing through the Use of Character Structure	201
<i>Yannis Haralambous</i>	
The Application of Kalman Filter Based Human-Computer Learning Model to Chinese Word Segmentation	218
<i>Weimeng Zhu, Ni Sun, Xiaojun Zou, and Junfeng Hu</i>	
Machine Learning for High-Quality Tokenization Replicating Variable Tokenization Schemes	231
<i>Murhaf Fares, Stephan Oepen, and Yi Zhang</i>	

Syntax and Named Entity Recognition

Structural Prediction in Incremental Dependency Parsing	245
<i>Niels Beuck and Wolfgang Menzel</i>	
Semi-supervised Constituent Grammar Induction Based on Text Chunking Information	258
<i>Jesús Santamaría and Lourdes Araújo</i>	
Turkish Constituent Chunking with Morphological and Contextual Features	270
<i>İlknur Durgar El-Kahlout and Ahmet Afşın Akın</i>	

Enhancing Czech Parsing with Verb Valency Frames	282
<i>Miloš Jakubíček and Vojtěch Kovář</i>	
An Automatic Approach to Treebank Error Detection Using a Dependency Parser	294
<i>Bhasha Agrawal, Rahul Agarwal, Samar Husain, and Dipti M. Sharma</i>	
Topic-Oriented Words as Features for Named Entity Recognition (Best Paper Award, Second Place)	304
<i>Ziqi Zhang, Trevor Cohn, and Fabio Ciravegna</i>	
Named Entities in Judicial Transcriptions: Extended Conditional Random Fields	317
<i>Elisabetta Fersini and Enza Messina</i>	
Introducing Baselines for Russian Named Entity Recognition	329
<i>Rinat Gareev, Maksim Tkachenko, Valery Solovyev, Andrey Simanovsky, and Vladimir Ivanov</i>	

Word Sense Disambiguation and Coreference Resolution

Five Languages Are Better Than One: An Attempt to Bypass the Data Acquisition Bottleneck for WSD (Best Paper Award, Third Place)	343
<i>Els Lefever, Véronique Hoste, and Martine De Cock</i>	
Analyzing the Sense Distribution of Concordances Obtained by Web as Corpus Approach	355
<i>Xabier Saralegi and Pablo Gamallo</i>	
MaxMax: A Graph-Based Soft Clustering Algorithm Applied to Word Sense Induction	368
<i>David Hope and Bill Keller</i>	
A Model of Word Similarity Based on Structural Alignment of Subject-Verb-Object Triples	382
<i>Dervla O’Keeffe and Fintan Costello</i>	
Coreference Annotation Schema for an Inflectional Language	394
<i>Maciej Ogrodniczuk, Magdalena Zawistawska, Katarzyna Głowińska, and Agata Savary</i>	
Exploring Coreference Uncertainty of Generically Extracted Event Mentions	408
<i>Goran Glavaš and Jan Šnajder</i>	

Semantics and Discourse

LIAR _C : Labeling Implicit ARguments in Spanish Deverbal Nominalizations	423
<i>Aina Peris, Mariona Taulé, Horacio Rodríguez, and Manuel Bertran Ibarz</i>	
Automatic Detection of Idiomatic Clauses (Best Paper Award, First Place)	435
<i>Anna Feldman and Jing Peng</i>	
Evaluating the Results of Methods for Computing Semantic Relatedness	447
<i>Felice Ferrara and Carlo Tasso</i>	
Similarity Measures Based on Latent Dirichlet Allocation	459
<i>Vasile Rus, Nobal Niraula, and Rajendra Banjade</i>	
Evaluating the Premises and Results of Four Metaphor Identification Systems	471
<i>Jonathan Dunn</i>	
Determining the Conceptual Space of Metaphoric Expressions	487
<i>David B. Bracewell, Marc T. Tomlinson, and Michael Mohler</i>	
What is being Measured in an Information Graphic?	501
<i>Seniz Demir, Stephanie Elzer Schwartz, Richard Burns, and Sandra Carberry</i>	
Comparing Discourse Tree Structures	513
<i>Elena Mitocariu, Daniel Alexandru Anechitei, and Dan Cristea</i>	
Assessment of Different Workflow Strategies for Annotating Discourse Relations: A Case Study with HDRB	523
<i>Himanshu Sharma, Praveen Dakwale, Dipti M. Sharma, Rashmi Prasad, and Aravind Joshi</i>	
Building a Discourse Parser for Informal Mathematical Discourse in the Context of a Controlled Natural Language	533
<i>Raúl Ernesto Gutiérrez de Piñerez Reyes and Juan Francisco Díaz Frias</i>	
Discriminative Learning of First-Order Weighted Abduction from Partial Discourse Explanations	545
<i>Kazeto Yamamoto, Naoya Inoue, Yotaro Watanabe, Naoaki Okazaki, and Kentaro Inui</i>	

Facilitating the Analysis of Discourse Phenomena in an Interoperable NLP Platform	559
<i>Rıza Theresa Batista-Navarro, Georgios Kontonatsios,</i> <i>Claudiu Mihăilă, Paul Thompson, Rafal Rak, Raheel Nawaz,</i> <i>Ioannis Korkontzelos, and Sophia Ananiadou</i>	
Author Index	573

Table of Contents – Part II

Sentiment, Polarity, Emotion, Subjectivity, and Opinion

Damping Sentiment Analysis in Online Communication: Discussions, Monologs and Dialogs (Invited paper)	1
<i>Mike Thelwall, Kevan Buckley, George Paltoglou, Marcin Skowron, David Garcia, Stephane Gobron, Junghyun Ahn, Arvid Kappas, Dennis Küster, and Janusz A. Holyst</i>	
Optimal Feature Selection for Sentiment Analysis	13
<i>Basant Agarwal and Namita Mittal</i>	
Measuring the Effect of Discourse Structure on Sentiment Analysis	25
<i>Baptiste Chardon, Farah Benamara, Yannick Mathieu, Vladimir Popescu, and Nicholas Asher</i>	
Lost in Translation: Viability of Machine Translation for Cross Language Sentiment Analysis	38
<i>Balamurali A.R., Mitesh M. Khapra, and Pushpak Bhattacharyya</i>	
An Enhanced Semantic Tree Kernel for Sentiment Polarity Classification	50
<i>Luis A. Trindade, Hui Wang, William Blackburn, and Niall Rooney</i>	
Combining Supervised and Unsupervised Polarity Classification for non-English Reviews	63
<i>José M. Perea-Ortega, Eugenio Martínez-Cámara, María-Teresa Martín-Valdivia, and L. Alfonso Ureña-López</i>	
Word Polarity Detection Using a Multilingual Approach	75
<i>Cüneyd Murad Özsert and Arzucan Özgür</i>	
Mining Automatic Speech Transcripts for the Retrieval of Problematic Calls	83
<i>Frederik Cailliau and Ariane Cavet</i>	
Cross-Lingual Projections vs. Corpora Extracted Subjectivity Lexicons for Less-Resourced Languages	96
<i>Xabier Saralegi, Iñaki San Vicente, and Irati Ugarteburu</i>	
Predicting Subjectivity Orientation of Online Forum Threads	109
<i>Prakhar Biyani, Cornelia Caragea, and Prasenjit Mitra</i>	

Distant Supervision for Emotion Classification with Discrete Binary Values	121
<i>Jared Suttles and Nancy Ide</i>	
Using Google n-grams to Expand Word-Emotion Association Lexicon	137
<i>Jessica Perrie, Aminul Islam, Evangelos Milios, and Vlado Keselj</i>	
A Joint Prediction Model for Multiple Emotions Analysis in Sentences	149
<i>Yunong Wu, Kenji Kita, Kazuyuki Matsumoto, and Xin Kang</i>	
Evaluating the Impact of Syntax and Semantics on Emotion Recognition from Text	161
<i>Gözde Özbal and Daniele Pighin</i>	
Chinese Emotion Lexicon Developing via Multi-lingual Lexical Resources Integration	174
<i>Jun Xu, Ruifeng Xu, Yanzhen Zheng, Qin Lu, Kai-Fai Wong, and Xiaolong Wang</i>	
N-Gram-Based Recognition of Threatening Tweets	183
<i>Nelleke Oostdijk and Hans van Halteren</i>	
Distinguishing the Popularity between Topics: A System for Up-to-Date Opinion Retrieval and Mining in the Web	197
<i>Nikolaos Pappas, Georgios Katsimpras, and Efsthios Stamatatos</i>	

Machine Translation and Multilingualism

No Free Lunch in Factored Phrase-Based Machine Translation	210
<i>Aleš Tamchyna and Ondřej Bojar</i>	
Domain Adaptation in Statistical Machine Translation Using Comparable Corpora: Case Study for English Latvian IT Localisation (Best Student Paper Award)	224
<i>Mārcis Pinnis, Inguna Skadiņa, and Andrejs Vasiļjevs</i>	
Assessing the Accuracy of Discourse Connective Translations: Validation of an Automatic Metric	236
<i>Najeh Hajlaoui and Andrei Popescu-Belis</i>	
An Empirical Study on Word Segmentation for Chinese Machine Translation	248
<i>Hai Zhao, Masao Utiyama, Eiichiro Sumita, and Bao-Liang Lu</i>	
Class-Based Language Models for Chinese-English Parallel Corpus	264
<i>Junfei Guo, Juan Liu, Michael Walsh, and Helmut Schmid</i>	

Building a Bilingual Dictionary from a Japanese-Chinese Patent Corpus	276
<i>Keiji Yasuda and Eiichiro Sumita</i>	
A Diagnostic Evaluation Approach for English to Hindi MT Using Linguistic Checkpoints and Error Rates	285
<i>Renu Balyan, Sudip Kumar Naskar, Antonio Toral, and Niladri Chatterjee</i>	
Leveraging Arabic-English Bilingual Corpora with Crowd Sourcing-Based Annotation for Arabic-Hebrew SMT	297
<i>Manish Gaurav, Guruprasad Saikumar, Amit Srivastava, Premkumar Natarajan, Shankar Ananthakrishnan, and Spyros Matsoukas</i>	
Automatic and Human Evaluation on English-Croatian Legislative Test Set	311
<i>Marija Brkić, Sanja Seljan, and Tomislav Vičić</i>	
Text Mining, Information Extraction, and Information Retrieval	
Enhancing Search: Events and Their Discourse Context (Invited paper)	318
<i>Sophia Ananiadou, Paul Thompson, and Raheel Nawaz</i>	
Distributional Term Representations for Short-Text Categorization	335
<i>Juan Manuel Cabrera, Hugo Jair Escalante, and Manuel Montes-y-Gómez</i>	
Learning Bayesian Network Using Parse Trees for Extraction of Protein-Protein Interaction	347
<i>Pedro Nelson Shiguihara-Juárez and Alneu de Andrade Lopes</i>	
A Model for Information Extraction in Portuguese Based on Text Patterns	359
<i>Tiago Luis Bonamigo and Renata Vieira</i>	
A Study on Query Expansion Based on Topic Distributions of Retrieved Documents	369
<i>Midori Serizawa and Ichiro Kobayashi</i>	
Link Analysis for Representing and Retrieving Legal Information	380
<i>Alfredo López Monroy, Hiram Calvo, Alexander Gelbukh, and Georgina García Pacheco</i>	

Text Summarization

Discursive Sentence Compression 394
*Alejandro Molina, Juan-Manuel Torres-Moreno, Eric SanJuan,
Iria da Cunha, and Gerardo Eugenio Sierra Martínez*

A Knowledge Induced Graph-Theoretical Model for Extract and
Abstract Single Document Summarization 408
Niraj Kumar, Kannan Srinathan, and Vasudeva Varma

Hierarchical Clustering in Improving Microblog Stream
Summarization 424
Andrei Olariu

Summary Evaluation: Together We Stand NPower-ed 436
George Giannakopoulos and Vangelis Karkaletsis

Stylometry and Text Simplification

Explanation in Computational Stylometry (Invited paper) 451
Walter Daelemans

The Use of Orthogonal Similarity Relations in the Prediction of
Authorship 463
*Upendra Sapkota, Thamar Solorio, Manuel Montes-y-Gómez, and
Paolo Rosso*

ERNESTA: A Sentence Simplification Tool for Children’s Stories in
Italian 476
Gianni Barlacchi and Sara Tonelli

Automatic Text Simplification in Spanish: A Comparative Evaluation
of Complementing Modules 488
*Biljana Drndarević, Sanja Štajner, Stefan Bott,
Susana Bautista, and Horacio Saggion*

The Impact of Lexical Simplification by Verbal Paraphrases for People
with and without Dyslexia 501
Luz Rello, Ricardo Baeza-Yates, and Horacio Saggion

Detecting Apposition for Text Simplification in Basque 513
*Itziar Gonzalez-Dios, María Jesús Aranzabe,
Arantza Díaz de Ilarraza, and Ander Soraluze*

Applications

Automation of Linguistic Creativitas for Adslogia 525
Gözde Özbal and Carlo Strapparava

Allongos: Longitudinal Alignment for the Genetic Study of Writers’ Drafts	537
<i>Adrien Lardilleux, Serge Fleury, and Georgeta Cislaru</i>	
A Combined Method Based on Stochastic and Linguistic Paradigm for the Understanding of Arabic Spontaneous Utterances	549
<i>Chahira Lhioui, Anis Zouaghi, and Mounir Zrigui</i>	
Evidence in Automatic Error Correction Improves Learners’ English Skill	559
<i>Jiro Umezawa, Junta Mizuno, Naoaki Okazaki, and Kentaro Inui</i>	
Author Index	573