

EACL 2012

**13th Conference of the European Chapter of the
Association for Computational Linguistics**

Proceedings of the Conference

April 23 - 27 2012
Avignon France

© 2012 The Association for Computational Linguistics

ISBN 978-1-937284-19-0

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Preface: General Chair

Welcome to EACL 2012, the 13th Conference of the European Chapter of the Association for Computational Linguistics. We are happy that despite strong competition from other Computational Linguistics events and economic turmoil in many European countries, this EACL is comparable to the successful previous ones, both in terms of the number of papers submitted and in terms of attendance. We have a strong scientific program, including ten workshops, four tutorials, a demos session and a student research workshop. I am convinced that you will appreciate our program.

What does a General Chair at EACL have to do? Not much, it turns out. My job was to act as a liaison between the local organizing team, the scientific committees, and the EACL board, and to give advice when needed. Looking back at the thousands of e-mails I was copied on reminded me of the Jerome K. Jerome quote: "I like work. I can sit and look at it for hours". It has been an enjoyable experience to cooperate with the many people who made this conference happen, and to see them work. I have learned a lot from them.

The Program Committee at an ACL conference is a trained army of Area Chairs, Program Committee members, and additional reviewers. Mirella Lapata and Lluís Màrquez commanded this particular one. It is thanks to the voluntary peer reviewing work, year after year, of this large group of people, formed by the top researchers in our field, that you will find a high-quality program. It is thanks to Mirella and Lluís that you will not only find the quality we expect from EACL, but also innovation, coherence, breadth, and depth. I can't thank them enough for their work on all aspects of the scientific program and for their advice on virtually any other aspect of the organization. Many thanks also to Regina Barzilay, Raymond Mooney, and Martin Cooke for accepting to present an invited lecture and thereby increase the appeal of this event even more.

As in previous years, the selection of the workshops of all ACL conferences in the same year is coordinated in a single committee. For EACL, Kristiina Jokinen and Alessandro Moschitti collaborated with the NAACL and ACL chairs in reviewing and selecting the workshops. As EACL is the first conference of the three, they had to initiate the call for proposals and activate their colleagues long before they were planning to. Thanks to their professionalism and efficiency, the process went very smoothly, and the resulting workshops program reflects the diversity and maturity of the field. For even more variation during the first two days of the conference, we also have a strong tutorial program. Tutorial Chairs Lieve Macken and Eneko Agirre managed to attract an impressive list of high-quality submissions and performed a thorough and thoughtful review and selection. It is truly a pity only four could be accommodated in the program, but their quality and timeliness is inspiring. Many thanks to Kristiina, Alessandro, Lieve, and Eneko for making this important part of the scientific program such a success.

As in previous editions of EACL, the Student Research Workshop was organized by the student members of the EACL board: Pierre Lison, Mattias Nilsson, and Marta Recasens, with help from faculty advisor Laurence Danlos. Their task was a huge one: to organize a mini-conference within the conference. This included finding reviewers, selecting papers, setting up a program for the student session, finding mentors for the accepted papers, selecting a best paper award, ... The amount of work they did cannot be overestimated, and the result is brilliant. Thank you! To round of the scientific program, we have stimulating demonstration sessions, selected and coordinated by Demonstrations Chair Frédérique Segond. Thank you for showing so clearly the rapid progress application-oriented computational linguistics is making.

Thanks also to Gertjan van Noord and Caroline Sporleder for accepting the role of coordinators of the mentoring service. In the end, they didn't have to assign mentors, but it is important that such a service is available when needed.

For EACL 2012 we decided to switch to digital proceedings only. They were available before the conference from the website, during the conference on the memory stick you received with your registration material, and afterwards from the website and the ACL Anthology. An exception was made for the tutorial notes, which are available to participants on paper as well. I warned the Publications Chairs, Adrià de Gispert and Fabrice Lefèvre, beforehand that theirs was probably the most demanding and stressful task of the conference: making sure that huge volumes of material from so many sources are available in time and in the right format, incorporating last minute corrections, and handling unavoidable glitches in the publications software. It is a formidable task, but they completed it without flinching. We all owe them our gratitude.

EACL seems to follow economical crises, let us hope it does not become a habit. Both the previous conference in 2009 and the current one happened in grim economical times. Being a Sponsorship Chair is not a happy occasion in such times. Nevertheless, both the international ACL Sponsorship Committee (with Massimiliano Ciaramita as EACL member) and the local Sponsorship Chairs (Eric SanJuan and Stéphane Huet) left no stone unturned looking for sponsors. We would have ended up in a much worse financial situation if it hadn't been for their efforts. Thank you! And of course also many thanks to our sponsors who, despite the economic situation, decided to help us financially with the conference. I am convinced their investment will be rewarded.

Organizing large conferences like this is a complex undertaking, even with the help of extensive material (the ACL conference handbook). Whenever in doubt, I have had the opportunity to interact with the EACL Board, and occasionally with the ACL Board and with Priscilla Rasmussen. This has always been a pleasure. I have learned that the people running our associations are dedicated, know everything, and never sleep.

Last but not least, the local organizing team has had to carry the largest burden in the organization. The sheer number of tasks and actions the local organizers of a conference like EACL have to assume is astonishing. Marc El-Beze has been a wonderful chair and his team (Frederic Bechet, Yann Fernandez, Stéphane Huet, Tania Jimenez, Fabrice Lefevre, Georges Linares, Alexis Nasr, Eric SanJuan, and Iria Da Cunha) has done outstanding work. There is no beginning in mentioning the many tasks they had to fulfill for making this a top conference. I am very grateful for all the work they put in the event and for the stress-free and friendly cooperation. I am also grateful for the support of the University of Avignon.

I hope you will have many fond memories of EACL 2012, organized in these stunning surroundings in Avignon, both about the exciting scientific program and about the superb social program and local arrangements.

Walter Daelemans
General Chair
March 2012

Preface: Program Chairs

We are delighted to present you with this volume containing the papers accepted for presentation at the 13th Conference of the European Chapter of the Association for Computational Linguistics, held in Avignon, France, from April 23 till April 27 2012.

EACL 2012 received 326 submissions. We were able to accept 85 papers in total (an acceptance rate of 26%). 48 of the papers (14.7%) were accepted for oral presentation, and 34 (10.4%) for poster presentation. One oral paper was subsequently withdrawn after acceptance. The papers were selected by a program committee of 28 area chairs, from Asia, Europe, and North America, assisted by a panel of 471 reviewers. Each submission was reviewed by three reviewers, who were furthermore encouraged to discuss any divergences they might have, and the papers in each area were ranked by the area chairs. The final selection was made by the program co-chairs after an independent check of all reviews and discussions with the area chairs.

This year EACL introduced an author response period. Authors were able to read and respond to the reviews of their paper before the program committee made a final decision. They were asked to correct factual errors in the reviews and answer questions raised in the reviewers comments. The intention was to help produce more accurate reviews. In some cases, reviewers changed their scores in view of the authors response and the area chairs read all responses carefully prior to making recommendations for acceptance. Another new feature was to allow authors to include optional supplementary material in addition to the paper itself (e.g., code, data sets, and resources). Finally, in an attempt to eliminate any bias from the reviewing process we put in place a double-blind reviewing system where the identity of the authors was not revealed to the area chairs.

After the program was selected, each of the area chairs was asked to nominate the best paper from his or her area, or to explicitly decline to nominate any. This resulted in several nominations out of which three stood out and were further considered in more detail by an dedicated committee chaired by Stephen Clark. This independent committee selected the best paper of the conference, which will be also awarded with a prize sponsored by Google. The best paper and the other two finalists will be presented in plenary sessions at the conference.

In addition to the main conference program, EACL 2012 will feature the now traditional Student Research Workshop, 10 workshops, 4 tutorials and a demo session with 21 presentations. We are also fortunate to have three invited speakers, Martin Cooke, Ikerbasque (Basque Foundation for Science), Regina Barzilay, Massachusetts Institute of Technology, and Raymond Mooney, University of Texas at Austin. Martin Cooke will speak about “Speech Communication in the Wild”, Regina Barzilay will discuss the topic of “Learning to Behave by Reading”, and Raymond Mooney will present on “Learning Language from Perceptual Context”.

First and foremost, we would like to thank the authors who submitted their work to EACL. The sheer number of submissions reflects how broad and active our field is. We are deeply indebted to the area chairs and the reviewers for their hard work. They enabled us to select an exciting program and to provide valuable feedback to the authors. We are grateful to our invited speakers who graciously agreed to give talks at EACL. Additional thanks to the Publications Chairs, Adrià de Gispert and Fabrice Lefèvre who put this volume together. We are grateful to Rich Gerber and the START team who always responded to our questions quickly, and helped us manage the large number of submissions smoothly. Thanks are due to the local organizing committee chair, Marc El-Beze for his cooperation with us over many organisational issues. We are also grateful to the Student Research Workshop chairs, Pierre Lison, Mattias Nilsson, and Marta Recasens, and the NAACL-HLT (Srinivas Bangalore, Eric Fosler-Lussier and Ellen Riloff) and ACL (Chin-Yew Lin and Miles Osborne) program chairs for their smooth collaboration in the handling of double submissions. Last but not least, we are indebted to the

General Chair, Walter Daelemans, for his guidance and support throughout the whole process.

We hope you enjoy the conference!

Mirella Lapata and Lluís Márquez

EACL 2012 Program Chairs

Organizing Committee

General Chair:

Walter Daelemans, University of Antwerp, Belgium

Programme Committee Chairs:

Mirella Lapata, University of Edinburgh, UK

Lluís Màrquez, Universitat Politècnica de Catalunya, Spain

Area Chairs:

Katja Filippova, Google

Min-Yen Kan, National University of Singapore

Charles Sutton, University of Edinburgh

Ivan Titov, Saarland University

Xavier Carreras, Universitat Politècnica de Catalunya (UPC)

Kenji Sagae, University of Southern California

Kallirroi Georgila, Institute for Creative Technologies, University of Southern California

Michael Strube, HITS gGmbH

Pascale Fung, The Hong Kong University of Science and Technology

Bing Liu, University of Illinois at Chicago

Theresa Wilson, Johns Hopkins University

David McClosky, Stanford University

Sebastian Riedel, University of Massachusetts

Phil Blunsom, University of Oxford

Mikel L. Forcada, Universitat d'Alacant

Christof Monz, University of Amsterdam

Sharon Goldwater, University of Edinburgh

Richard Wicentowski, Swarthmore College

Patrick Pantel, Microsoft Research

Hiroya Takamura, Tokyo Institute of Technology

Alexander Koller, University of Potsdam

Sebastian Padó, Universität Heidelberg

Maarten de Rijke, University of Amsterdam

Julio Gonzalo, UNED

Lori Levin, Carnegie Mellon University

Piek Vossen, VU University Amsterdam

Afra Alishahi, Tilburg University, The Netherlands

John Hale, Cornell University

Workshop Committee chairs:

Kristiina Jokinen, University of Helsinki, Finland

Alessandro Moschitti, University of Trento, Italy

Tutorials Committee chairs:

Eneko Agirre, University of the Basque Country, Spain

Lieve Macken, University College Ghent, Belgium

Student research workshop chairs:

Pierre Lison, University of Oslo, Norway
Mattias Nilsson, Uppsala University, Sweden
Marta Recasens, University of Barcelona, Spain

Student research workshop faculty advisor:

Laurence Danlos, Université Paris 7

System Demonstrations Committee:

Frédérique Segond

Publications Committee:

Adrià de Gispert, University of Cambridge, UK
Fabrice Lefèvre, University of Avignon, France

Sponsorship Committee:

Massimiliano Ciaramita

Mentoring service:

Caroline Sporleder, Saarland University, Germany
Gertjan van Noord, University of Groningen, The Netherlands

Local Organising Committee:

Marc El-Beze (Chair), University of Avignon, France
Frederic Bechet (Publicity chair), University Aix-Marseille 2, France
Yann Fernandez, University of Avignon, France
Stéphane Huet (Exhibits local chair), University of Avignon, France
Tania Jimenez, University of Avignon, France
Fabrice Lefevre, University of Avignon, France
Georges Linares, University of Avignon, France
Alexis Nasr, University Aix-Marseille 2, France
Eric SanJuan (Sponsorship local chair), University of Avignon, France
Iria Da Cunha, Pompeu Fabra University, Spain

Table of Contents

<i>Speech Communication in the Wild</i>	
Martin Cooke	1
<i>Power-Law Distributions for Paraphrases Extracted from Bilingual Corpora</i>	
Spyros Martzoukos and Christof Monz	2
<i>A Bayesian Approach to Unsupervised Semantic Role Induction</i>	
Ivan Titov and Alexandre Klementiev	12
<i>Entailment above the word level in distributional semantics</i>	
Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do and Chung-chieh Shan	23
<i>Evaluating Distributional Models of Semantics for Syntactically Invariant Inference</i>	
Jackie Chi Kit Cheung and Gerald Penn	33
<i>Cross-Framework Evaluation for Statistical Parsing</i>	
Reut Tsarfaty, Joakim Nivre and Evelina Andersson	44
<i>Dependency Parsing of Hungarian: Baseline Results and Challenges</i>	
Richárd Farkas, Veronika Vincze and Helmut Schmid	55
<i>Dependency Parsing with Undirected Graphs</i>	
Carlos Gómez-Rodríguez and Daniel Fernández-González	66
<i>The Best of Both Worlds – A Graph-based Completion Model for Transition-based Parsers</i>	
Bernd Bohnet and Jonas Kuhn	77
<i>Answer Sentence Retrieval by Matching Dependency Paths acquired from Question/Answer Sentence Pairs</i>	
Michael Kaisser	88
<i>Can Click Patterns across User’s Query Logs Predict Answers to Definition Questions?</i>	
Alejandro Figueroa	99
<i>Adaptation of Statistical Machine Translation Model for Cross-Lingual Information Retrieval in a Service Context</i>	
Vassilina Nikoulina, Bogomil Kovachev, Nikolaos Lagos and Christof Monz	109
<i>Computing Lattice BLEU Oracle Scores for Machine Translation</i>	
Artem Sokolov, Guillaume Wisniewski and Francois Yvon	120
<i>Toward Statistical Machine Translation without Parallel Corpora</i>	
Alexandre Klementiev, Ann Irvine, Chris Callison-Burch and David Yarowsky	130
<i>Character-Based Pivot Translation for Under-Resourced Languages and Domains</i>	
Jörg Tiedemann	141
<i>Does more data always yield better translations?</i>	
Guillem Gascó, Martha-Alicia Rocha, Germán Sanchis-Trilles, Jesús Andrés-Ferrer and Francisco Casacuberta	152
<i>Recall-Oriented Learning of Named Entities in Arabic Wikipedia</i>	
Behrang Mohit, Nathan Schneider, Rishav Bhowmick, Kemal Oflazer and Noah A. Smith	162

<i>Tree Representations in Probabilistic Models for Extended Named Entities Detection</i> Marco Dinarelli and Sophie Rosset	174
<i>When Did that Happen? — Linking Events and Relations to Timestamps</i> Dirk Hovy, James Fan, Alfio Gliozzo, Siddharth Patwardhan and Christopher Welty	185
<i>Compensating for Annotation Errors in Training a Relation Extractor</i> Bonan Min and Ralph Grishman	194
<i>Incorporating Lexical Priors into Topic Models</i> Jagadeesh Jagarlamudi, Hal Daume III and Raghavendra Udupa	204
<i>DualSum: a Topic-Model based approach for update summarization</i> Jean-Yves Delort and Enrique Alfonseca	214
<i>Large-Margin Learning of Submodular Summarization Models</i> Ruben Sipos, Pannaga Shivaswamy and Thorsten Joachims	224
<i>A Probabilistic Model of Syntactic and Semantic Acquisition from Child-Directed Utterances and their Meanings</i> Tom Kwiatkowski, Sharon Goldwater, Luke Zettlemoyer and Mark Steedman	234
<i>Active learning for interactive machine translation</i> Jesús González-Rubio, Daniel Ortiz-Martínez and Francisco Casacuberta	245
<i>Adapting Translation Models to Translationese Improves SMT</i> Gennadi Lembersky, Noam Ordan and Shuly Wintner	255
<i>Aspectual Type and Temporal Relation Classification</i> Francisco Costa and António Branco	266
<i>Automatic generation of short informative sentiment summaries</i> Andrea Glaser and Hinrich Schütze	276
<i>Bootstrapped Training of Event Extraction Classifiers</i> Ruihong Huang and Ellen Riloff	286
<i>Bootstrapping Events and Relations from Text</i> Ting Liu and Tomek Strzalkowski	296
<i>CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language</i> Svitlana Volkova, William B. Dolan and Theresa Wilson	306
<i>Extending the Entity-based Coherence Model with Multiple Ranks</i> Vanessa Wei Feng and Graeme Hirst	315
<i>Generalization Methods for In-Domain and Cross-Domain Opinion Holder Extraction</i> Michael Wiegand and Dietrich Klakow	325
<i>Skip N-grams and Ranking Functions for Predicting Script Events</i> Bram Jans, Steven Bethard, Ivan Vulić and Marie-Francine Moens	336
<i>The Problem with Kappa</i> David Martin Ward Powers	345

<i>User Edits Classification Using Document Revision Histories</i> Amit Bronner and Christof Monz	356
<i>User Participation Prediction in Online Forums</i> Zhonghua Qu and Yang Liu	367
<i>Inferring Selectional Preferences from Part-Of-Speech N-grams</i> Hyeju Jang and Jack Mostow	377
<i>WebCAGe – A Web-Harvested Corpus Annotated with GermaNet Senses</i> Verena Henrich, Erhard Hinrichs and Tatiana Vodolazova	387
<i>Learning to Behave by Reading</i> Regina Barzilay	397
<i>Lexical surprisal as a general predictor of reading time</i> Irene Fernandez Monsalve, Stefan L. Frank and Gabriella Vigliocco	398
<i>Spectral Learning for Non-Deterministic Dependency Parsing</i> Franco M. Luque, Ariadna Quattoni, Borja Balle and Xavier Carreras	409
<i>Combining Tree Structures, Flat Features and Patterns for Biomedical Relation Extraction</i> Md. Faisal Mahbub Chowdhury and Alberto Lavelli	420
<i>Coordination Structure Analysis using Dual Decomposition</i> Atsushi Hanamoto, Takuya Matsuzaki and Jun'ichi Tsujii	430
<i>Cutting the Long Tail: Hybrid Language Models for Translation Style Adaptation</i> Arianna Bisazza and Marcello Federico	439
<i>Detecting Highly Confident Word Translations from Comparable Corpora without Any Prior Knowledge</i> Ivan Vulić and Marie-Francine Moens	449
<i>Efficient parsing with Linear Context-Free Rewriting Systems</i> Andreas van Cranenburgh	460
<i>Evaluating language understanding accuracy with respect to objective outcomes in a dialogue system</i> Myroslava O. Dzikovska, Peter Bell, Amy Isard and Johanna D. Moore	471
<i>Experimenting with Distant Supervision for Emotion Classification</i> Matthew Purver and Stuart Battersby	482
<i>Feature-Rich Part-of-speech Tagging for Morphologically Complex Languages: Application to Bulgarian</i> Georgi Georgiev, Valentin Zhikov, Kiril Simov, Petya Osenova and Preslav Nakov	492
<i>Instance-Driven Attachment of Semantic Annotations over Conceptual Hierarchies</i> Janara Christensen and Marius Pasca	503
<i>Joint Satisfaction of Syntactic and Pragmatic Constraints Improves Incremental Spoken Language Understanding</i> Andreas Peldszus, Okko Buß, Timo Baumann and David Schlangen	514
<i>Learning How to Conjugate the Romanian Verb. Rules for Regular and Partially Irregular Verbs</i> Liviu P. Dinu, Vlad Niculae and Octavia-Maria Sulea	524

<i>Measuring Contextual Fitness Using Error Contexts Extracted from the Wikipedia Revision History</i> Torsten Zesch	529
<i>Perplexity Minimization for Translation Model Domain Adaptation in Statistical Machine Translation</i> Rico Sennrich	539
<i>Subcat-LMF: Fleshing out a standardized format for subcategorization frame interoperability</i> Judith Eckle-Kohler and Iryna Gurevych	550
<i>The effect of domain and text type on text prediction quality</i> Suzan Verberne, Antal van den Bosch, Helmer Strik and Lou Boves	561
<i>The Impact of Spelling Errors on Patent Search</i> Benno Stein, Dennis Hoppe and Tim Gollub	570
<i>UBY - A Large-Scale Unified Lexical-Semantic Resource Based on LMF</i> Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer and Christian Wirth	580
<i>Word Sense Induction for Novel Sense Detection</i> Jey Han Lau, Paul Cook, Diana McCarthy, David Newman and Timothy Baldwin	591
<i>Learning Language from Perceptual Context</i> Raymond Mooney	602
<i>Learning for Microblogs with Distant Supervision: Political Forecasting with Twitter</i> Micol Marchetti-Bowick and Nathanael Chambers	603
<i>Learning from evolving data streams: online triage of bug reports</i> Grzegorz Chrupala	613
<i>Towards a model of formal and informal address in English</i> Manaal Faruqui and Sebastian Pado	623
<i>Character-based kernels for novelistic plot structure</i> Micha Elsner	634
<i>Smart Paradigms and the Predictability and Complexity of Inflectional Morphology</i> Grégoire Détrez and Aarne Ranta	645
<i>Probabilistic Hierarchical Clustering of Morphological Paradigms</i> Burcu Can and Suresh Manandhar	654
<i>Modeling Inflection and Word-Formation in SMT</i> Alexander Fraser, Marion Weller, Aoife Cahill and Fabienne Cap	664
<i>Identifying Broken Plurals, Irregular Gender, and Rationality in Arabic Text</i> Sarah Alkuhlani and Nizar Habash	675
<i>Framework of Semantic Role Assignment based on Extended Lexical Conceptual Structure: Comparison with VerbNet and FrameNet</i> Yuichiroh Matsubayashi, Yusuke Miyao and Akiko Aizawa	686
<i>Unsupervised Detection of Downward-Entailing Operators By Maximizing Classification Certainty</i> Jackie Chi Kit Cheung and Gerald Penn	696

<i>Elliphant: Improved Automatic Detection of Zero Subjects and Impersonal Constructions in Spanish</i> Luz Rello, Ricardo Baeza-Yates and Ruslan Mitkov	706
<i>Validation of sub-sentential paraphrases acquired from parallel monolingual corpora</i> Houda Bouamor, Aurélien Max and Anne Vilnat	716
<i>Determining the placement of German verbs in English-to-German SMT</i> Anita Gojun and Alexander Fraser	726
<i>Syntax-Based Word Ordering Incorporating a Large-Scale Language Model</i> Yue Zhang, Graeme Blackwood and Stephen Clark	736
<i>Midge: Generating Image Descriptions From Computer Vision Detections</i> Margaret Mitchell, Jesse Dodge, Amit Goyal, Kota Yamaguchi, Karl Stratos, Xufeng Han, Alyssa Mensch, Alex Berg, Tamara Berg and Hal Daume III	747
<i>Generation of landmark-based navigation instructions from open-source data</i> Markus Dräger and Alexander Koller	757
<i>To what extent does sentence-internal realisation reflect discourse context? A study on word order</i> Sina Zarriß, Aoife Cahill and Jonas Kuhn	767
<i>Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages</i> Oliver Ferschke, Iryna Gurevych and Yevgen Chebotar	777
<i>An Unsupervised Dynamic Bayesian Network Approach to Measuring Speech Style Accommodation</i> Mahaveer Jain, John McDonough, Gahgene Gweon, Bhiksha Raj and Carolyn Penstein Rosé .	787
<i>Learning the Fine-Grained Information Status of Discourse Entities</i> Altaf Rahman and Vincent Ng	798
<i>Composing extended top-down tree transducers</i> Aurelie Lagoutte, Fabienne Braune, Daniel Quernheim and Andreas Maletti	808
<i>Structural and Topical Dimensions in Multi-Task Patent Translation</i> Katharina Waeschle and Stefan Riezler	818
<i>Not as Awful as it Seems: Explaining German Case through Computational Experiments in Fluid Construction Grammar</i> Remi van Trijp	829
<i>Managing Uncertainty in Semantic Tagging</i> Silvie Cinková, Martin Holub and Vincent Kríž	840
<i>Parallel and Nested Decomposition for Factoid Questions</i> Aditya Kalyanpur, Siddharth Patwardhan, Branimir Boguraev, Jennifer Chu-Carroll and Adam Lally	851

Conference Program

Wednesday April 25, 2012

(8:45) Session 1: Plenary Session

9:00 *Speech Communication in the Wild*
Martin Cooke

(10:30) Session 2a: Semantics

10:30 *Power-Law Distributions for Paraphrases Extracted from Bilingual Corpora*
Spyros Martzoukos and Christof Monz

10:55 *A Bayesian Approach to Unsupervised Semantic Role Induction*
Ivan Titov and Alexandre Klementiev

11:20 *Entailment above the word level in distributional semantics*
Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do and Chung-chieh Shan

11:45 *Evaluating Distributional Models of Semantics for Syntactically Invariant Inference*
Jackie Chi Kit Cheung and Gerald Penn

(10:30) Session 2b: Parsing

10:30 *Cross-Framework Evaluation for Statistical Parsing*
Reut Tsarfaty, Joakim Nivre and Evelina Andersson

10:55 *Dependency Parsing of Hungarian: Baseline Results and Challenges*
Richárd Farkas, Veronika Vincze and Helmut Schmid

11:20 *Dependency Parsing with Undirected Graphs*
Carlos Gómez-Rodríguez and Daniel Fernández-González

11:45 *The Best of Both Worlds – A Graph-based Completion Model for Transition-based Parsers*
Bernd Bohnet and Jonas Kuhn

Wednesday April 25, 2012 (continued)

(10:30) Session 2c: QA and IR

- 10:30 *Answer Sentence Retrieval by Matching Dependency Paths acquired from Question/Answer Sentence Pairs*
Michael Kaisser
- 10:55 *Can Click Patterns across User's Query Logs Predict Answers to Definition Questions?*
Alejandro Figueroa
- 11:20 *Adaptation of Statistical Machine Translation Model for Cross-Lingual Information Retrieval in a Service Context*
Vassilina Nikoulina, Bogomil Kovachev, Nikolaos Lagos and Christof Monz

(14:00) Session 3a: Machine Translation

- 14:00 *Computing Lattice BLEU Oracle Scores for Machine Translation*
Artem Sokolov, Guillaume Wisniewski and Francois Yvon
- 14:25 *Toward Statistical Machine Translation without Parallel Corpora*
Alexandre Klementiev, Ann Irvine, Chris Callison-Burch and David Yarowsky
- 14:50 *Character-Based Pivot Translation for Under-Resourced Languages and Domains*
Jörg Tiedemann
- 15:15 *Does more data always yield better translations?*
Guillem Gascó, Martha-Alicia Rocha, Germán Sanchis-Trilles, Jesús Andrés-Ferrer and Francisco Casacuberta

(14:00) Session 3b: Information Extraction

- 14:00 *Recall-Oriented Learning of Named Entities in Arabic Wikipedia*
Behrang Mohit, Nathan Schneider, Rishav Bhowmick, Kemal Oflazer and Noah A. Smith
- 14:25 *Tree Representations in Probabilistic Models for Extended Named Entities Detection*
Marco Dinarelli and Sophie Rosset
- 14:50 *When Did that Happen? — Linking Events and Relations to Timestamps*
Dirk Hovy, James Fan, Alfio Gliozzo, Siddharth Patwardhan and Christopher Welty

Wednesday April 25, 2012 (continued)

15:15 *Compensating for Annotation Errors in Training a Relation Extractor*
Bonan Min and Ralph Grishman

(14:00) Session 3c: Machine Learning and Summarization

14:00 *Incorporating Lexical Priors into Topic Models*
Jagadeesh Jagarlamudi, Hal Daume III and Raghavendra Udapa

14:25 *DualSum: a Topic-Model based approach for update summarization*
Jean-Yves Delort and Enrique Alfonseca

14:50 *Large-Margin Learning of Submodular Summarization Models*
Ruben Sipos, Pannaga Shivaswamy and Thorsten Joachims

(16:10) Session 4: Posters (1) and Demos (1)

16:10 *A Probabilistic Model of Syntactic and Semantic Acquisition from Child-Directed Utterances and their Meanings*
Tom Kwiatkowski, Sharon Goldwater, Luke Zettlemoyer and Mark Steedman

16:10 *Active learning for interactive machine translation*
Jesús González-Rubio, Daniel Ortiz-Martínez and Francisco Casacuberta

16:10 *Adapting Translation Models to Translationese Improves SMT*
Gennadi Lembersky, Noam Ordan and Shuly Wintner

16:10 *Aspectual Type and Temporal Relation Classification*
Francisco Costa and António Branco

16:10 *Automatic generation of short informative sentiment summaries*
Andrea Glaser and Hinrich Schütze

16:10 *Bootstrapped Training of Event Extraction Classifiers*
Ruihong Huang and Ellen Riloff

16:10 *Bootstrapping Events and Relations from Text*
Ting Liu and Tomek Strzalkowski

Wednesday April 25, 2012 (continued)

- 16:10 *CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language*
Svitlana Volkova, William B. Dolan and Theresa Wilson
- 16:10 *Extending the Entity-based Coherence Model with Multiple Ranks*
Vanessa Wei Feng and Graeme Hirst
- 16:10 *Generalization Methods for In-Domain and Cross-Domain Opinion Holder Extraction*
Michael Wiegand and Dietrich Klakow
- 16:10 *Skip N-grams and Ranking Functions for Predicting Script Events*
Bram Jans, Steven Bethard, Ivan Vulić and Marie-Francine Moens
- 16:10 *The Problem with Kappa*
David Martin Ward Powers
- 16:10 *User Edits Classification Using Document Revision Histories*
Amit Bronner and Christof Monz
- 16:10 *User Participation Prediction in Online Forums*
Zhonghua Qu and Yang Liu
- 16:10 *Inferring Selectional Preferences from Part-Of-Speech N-grams*
Hyeju Jang and Jack Mostow
- 16:10 *WebCAGe – A Web-Harvested Corpus Annotated with GermaNet Senses*
Verena Henrich, Erhard Hinrichs and Tatiana Vodolazova

Thursday April 26, 2012

(9:00) Session 5: Plenary Session

9:00 *Learning to Behave by Reading*
Regina Barzilay

(10:30) Session 6a: Student Workshop

(10:30) Session 6b: Student Workshop

(10:30) Session 6c: Student Workshop

(14:00) Session 7: EACL business meeting

(14:50) Session 8: Plenary Session

14:50 *Lexical surprisal as a general predictor of reading time*
Irene Fernandez Monsalve, Stefan L. Frank and Gabriella Vigliocco

15:15 *Spectral Learning for Non-Deterministic Dependency Parsing*
Franco M. Luque, Ariadna Quattoni, Borja Balle and Xavier Carreras

(16:10) Session 9: Posters (2) and Demos (2)

16:10 *Combining Tree Structures, Flat Features and Patterns for Biomedical Relation Extraction*
Md. Faisal Mahbub Chowdhury and Alberto Lavelli

16:10 *Coordination Structure Analysis using Dual Decomposition*
Atsushi Hanamoto, Takuya Matsuzaki and Jun'ichi Tsujii

16:10 *Cutting the Long Tail: Hybrid Language Models for Translation Style Adaptation*
Arianna Bisazza and Marcello Federico

16:10 *Detecting Highly Confident Word Translations from Comparable Corpora without Any Prior Knowledge*
Ivan Vulić and Marie-Francine Moens

Thursday April 26, 2012 (continued)

- 16:10 *Efficient parsing with Linear Context-Free Rewriting Systems*
Andreas van Cranenburgh
- 16:10 *Evaluating language understanding accuracy with respect to objective outcomes in a dialogue system*
Myroslava O. Dzikovska, Peter Bell, Amy Isard and Johanna D. Moore
- 16:10 *Experimenting with Distant Supervision for Emotion Classification*
Matthew Purver and Stuart Battersby
- 16:10 *Feature-Rich Part-of-speech Tagging for Morphologically Complex Languages: Application to Bulgarian*
Georgi Georgiev, Valentin Zhikov, Kiril Simov, Petya Osenova and Preslav Nakov
- 16:10 *Instance-Driven Attachment of Semantic Annotations over Conceptual Hierarchies*
Janara Christensen and Marius Pasca
- 16:10 *Joint Satisfaction of Syntactic and Pragmatic Constraints Improves Incremental Spoken Language Understanding*
Andreas Peldszus, Okko Buß, Timo Baumann and David Schlangen
- 16:10 *Learning How to Conjugate the Romanian Verb. Rules for Regular and Partially Irregular Verbs*
Liviu P. Dinu, Vlad Niculae and Octavia-Maria Sulea
- 16:10 *Measuring Contextual Fitness Using Error Contexts Extracted from the Wikipedia Revision History*
Torsten Zesch
- 16:10 *Perplexity Minimization for Translation Model Domain Adaptation in Statistical Machine Translation*
Rico Sennrich
- 16:10 *Subcat-LMF: Fleshing out a standardized format for subcategorization frame interoperability*
Judith Eckle-Kohler and Iryna Gurevych
- 16:10 *The effect of domain and text type on text prediction quality*
Suzan Verberne, Antal van den Bosch, Helmer Strik and Lou Boves
- 16:10 *The Impact of Spelling Errors on Patent Search*
Benno Stein, Dennis Hoppe and Tim Gollub

Thursday April 26, 2012 (continued)

- 16:10 *UBY - A Large-Scale Unified Lexical-Semantic Resource Based on LMF*
Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer and Christian Wirth
- 16:10 *Word Sense Induction for Novel Sense Detection*
Jey Han Lau, Paul Cook, Diana McCarthy, David Newman and Timothy Baldwin

Friday April 27, 2012

(9:00) Session 10: Plenary Session

- 9:00 *Learning Language from Perceptual Context*
Raymond Mooney

(10:30) Session 11a: Data Mining and Discourse

- 10:30 *Learning for Microblogs with Distant Supervision: Political Forecasting with Twitter*
Micol Marchetti-Bowick and Nathanael Chambers
- 10:55 *Learning from evolving data streams: online triage of bug reports*
Grzegorz Chrupala
- 11:20 *Towards a model of formal and informal address in English*
Manaal Faruqui and Sebastian Pado
- 11:45 *Character-based kernels for novelistic plot structure*
Micha Elsner

Friday April 27, 2012 (continued)

(10:30) Session 11b: Morphology

- 10:30 *Smart Paradigms and the Predictability and Complexity of Inflectional Morphology*
Grégoire Détrez and Aarne Ranta
- 10:55 *Probabilistic Hierarchical Clustering of Morphological Paradigms*
Burcu Can and Suresh Manandhar
- 11:20 *Modeling Inflection and Word-Formation in SMT*
Alexander Fraser, Marion Weller, Aoife Cahill and Fabienne Cap
- 11:45 *Identifying Broken Plurals, Irregular Gender, and Rationality in Arabic Text*
Sarah Alkuhlani and Nizar Habash

(10:30) Session 11c: Semantics

- 10:30 *Framework of Semantic Role Assignment based on Extended Lexical Conceptual Structure: Comparison with VerbNet and FrameNet*
Yuichiroh Matsubayashi, Yusuke Miyao and Akiko Aizawa
- 10:55 *Unsupervised Detection of Downward-Entailing Operators By Maximizing Classification Certainty*
Jackie Chi Kit Cheung and Gerald Penn
- 11:20 *Elliphant: Improved Automatic Detection of Zero Subjects and Impersonal Constructions in Spanish*
Luz Rello, Ricardo Baeza-Yates and Ruslan Mitkov
- 11:45 *Validation of sub-sentential paraphrases acquired from parallel monolingual corpora*
Houda Bouamor, Aurélien Max and Anne Vilnat

Friday April 27, 2012 (continued)

(14:00) Session 12a: Generation and Word Ordering

- 14:00 *Determining the placement of German verbs in English-to-German SMT*
Anita Gojun and Alexander Fraser
- 14:25 *Syntax-Based Word Ordering Incorporating a Large-Scale Language Model*
Yue Zhang, Graeme Blackwood and Stephen Clark
- 14:50 *Midge: Generating Image Descriptions From Computer Vision Detections*
Margaret Mitchell, Jesse Dodge, Amit Goyal, Kota Yamaguchi, Karl Stratos, Xufeng Han, Alyssa Mensch, Alex Berg, Tamara Berg and Hal Daume III
- 15:15 *Generation of landmark-based navigation instructions from open-source data*
Markus Dräger and Alexander Koller

(14:00) Session 12b: Discourse and Dialogue

- 14:00 *To what extent does sentence-internal realisation reflect discourse context? A study on word order*
Sina Zarriß, Aoife Cahill and Jonas Kuhn
- 14:25 *Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages*
Oliver Ferschke, Iryna Gurevych and Yevgen Chebotar
- 14:50 *An Unsupervised Dynamic Bayesian Network Approach to Measuring Speech Style Accommodation*
Mahaveer Jain, John McDonough, Gahgene Gweon, Bhiksha Raj and Carolyn Penstein Rosé
- 15:15 *Learning the Fine-Grained Information Status of Discourse Entities*
Altaf Rahman and Vincent Ng

Friday April 27, 2012 (continued)

(14:00) Session 12c: Parsing and MT

- 14:00 *Composing extended top-down tree transducers*
Aurelie Lagoutte, Fabienne Braune, Daniel Quernheim and Andreas Maletti
- 14:25 *Structural and Topical Dimensions in Multi-Task Patent Translation*
Katharina Waeschle and Stefan Riezler
- 14:50 *Not as Awful as it Seems: Explaining German Case through Computational Experiments in Fluid Construction Grammar*
Remi van Trijp

(15:45) Session 13: Plenary Session

- 15:45 *Managing Uncertainty in Semantic Tagging*
Silvie Cinková, Martin Holub and Vincent Kríž
- Parallel and Nested Decomposition for Factoid Questions*
Aditya Kalyanpur, Siddharth Patwardhan, Branimir Boguraev, Jennifer Chu-Carroll and Adam Lally