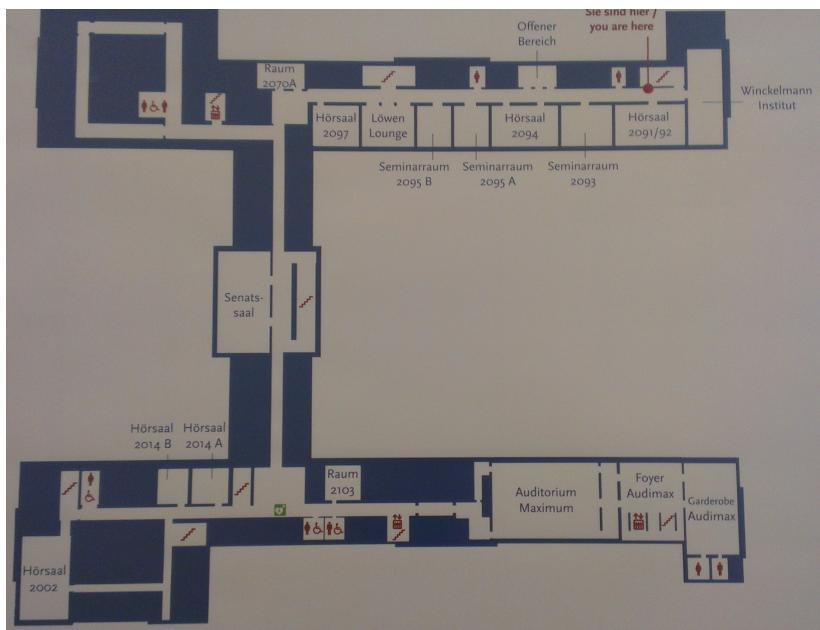
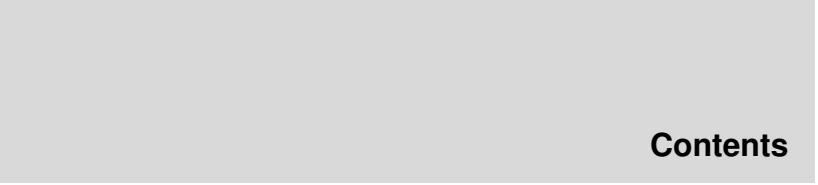


The 14th Conference of the
**North American Chapter of the Association for
Computational Linguistics:**
Human Language Technologies

CONFERENCE HANDBOOK



Cover design by ...
Handbook assembled by Florian Kunneman and Matt Post
Printing by Omnipress of Madison, Wisconsin



Contents

Conference Information

Message from the General Chair

Welcome to Berlin, welcome to ACL-2016!

A special welcome if this is your first ACL - I hope it fulfills your expectations and remains in your memory as a great start. Its magnitude may be a bit overwhelming - our field is on an expanding trajectory, and even a selection of the best work fills a great number of parallel sessions over a number of days; plus, there are the workshops to quench your topical thirst. Just dive in and drink it up.

If it is your first ACL, then take a good look occasionally at the people in your sessions, listening to your talk or grilling you at your poster; many of these people you will get to know over the years as you grow with them and go to the same meetings in the future. The ACL is not only a breeding ground for project ideas and future collaborations, but is also a hunting ground where job seekers find potential employers. Our new Recruitment Lunch was set up to foster this. Last but not least, ACLs have proven to be the starting moment for lasting friendships. Make the most of it!

To all others, welcome back to another great edition of ACL. The programme chairs and the chair teams for the demos, tutorials, and workshops have put together an exciting programme; the local chairs have fitted this all masterfully in Humboldt University's buildings at Unter den Linden in the very heart of Berlin. Cast in a quote in the marble entry hall, Karl Marx encourages us to use our insights and technologies to make the world a better place. This tantalizing encouragement is worth pondering over.

On behalf of the entire ACL-2016 organizing team, welcome to Berlin; I wish you a fruitful conference!

Antal van den Bosch
General Chair

Message from the Program Committee Co-Chairs

Welcome to the 54th Annual Meeting of the Association for Computational Linguistics! This year, ACL received 825 long paper submissions (a new record) and 463 short paper submissions.¹ Of the long papers, 231 were accepted for presentation at ACL—116 as oral presentations and 115 as poster presentations. 97 short papers were accepted—50 as oral and 47 as poster presentations. In addition, ACL also features 25 presentations of papers accepted in the *Transactions of the Association for Computational Linguistics* (TACL). With 353 papers being presented, this is the largest ACL program to date.

In keeping with the tremendous growth of our field, we introduced some changes to the conference. First, oral presentations were shortened to fifteen (twelve) minutes for long (short) papers, plus time for questions. While this places a greater demand on speakers to be concise, we believe it is worth the effort, allowing far more work to be presented orally. Second, we took advantage of the many halls available at Humboldt University and expanded the number of parallel talks during some conference sessions.

We introduced a category of outstanding papers to help recognize the highest quality work in the community this year. The 11 outstanding papers (9 long, 2 short, 0.85% of submissions) represent a broad spectrum of exciting contributions; they are recognized by especially prominent placement in the program. From these, a best paper and an IBM-sponsored best student paper have been selected; those will be announced in the awards session on Wednesday afternoon.

Following other recent ACL conferences, submissions were reviewed under different categories and using different review forms for empirical/data-driven, theoretical, applications/tools, resources/evaluation, and survey papers. We introduced special fields in the paper submission form for authors to explicitly note the release of open-source implementations to enable reproducibility, and to note freely available datasets. We also allowed authors to submit appendices of arbitrary length for details that would enable replication; reviewers were not expected to read this material.

Another innovation we explored during the review period was the scheduling of short paper review before long paper review. While this was planned to make the entire review period more compact (fitting between the constraints of NAACL 2016 and EMNLP 2016 at either end), we found that reviewing short papers first eliminated many of the surprises for the long paper review process.

We sought to follow recently-evolved best practices in planning the poster sessions, so that the many high-quality works presented in that format will be visible and authors and attendees benefit from the interactions during the two poster sessions.

ACL 2016 will have two distinguished invited speakers: Amber Boydston (Associate Professor of Political Science at the University of California, Davis) and Mark Steedman (Professor of Cognitive Science at the University of Edinburgh). We are grateful that they accepted our invitation and look forward to their presentations.

There are many individuals we wish to thank for their contributions to ACL 2016, some multiple times:

- The 38 area chairs who recruited reviewers, led the discussion about each paper, carefully assessed each submission, and authored meta-reviews to guide final decisions: Miguel Ballesteros, David Bamman, Steven Bethard, Jonathan Berant, Gemma Boleda, Ming-Wei Chang, Wanxiang Che, Chris Dyer, Ed Grefenstette, Hannaneh Hajishirzi, Minlie Huang, Mans Hulden, Heng Ji, Jing Jiang, Zornitsa Kozareva, Marco Kuhlmann, Yang Liu, Annie Louis, Wei Lu, Marie-Catherine de Marneffe, Gerard de Melo, David Mimno, Meg Mitchell, Daichi Mochihashi, Graham Neubig, Naoaki Okazaki, Simone Ponzetto,

¹These numbers exclude papers that were not reviewed due to formatting violations or that were withdrawn prior to review.

Matthew Purver, David Reitter, Nathan Schneider, Hinrich Schuetze, Thamar Solorio, Lucia Specia, Partha Talukdar, Ivan Titov, Lu Wang, Nianwen Xue, and Grace Yang.

- Our full program committee of 884 hard-working individuals who reviewed the conference's 1,288 submissions (including secondary reviewers).
- The ACL coordinating committee members, especially Yejin Choi, Graeme Hirst, Chris Manning, and Shiqi Zhao, who answered many questions as they arose during the year.
- TACL editors-in-chief Mark Johnson, Lillian Lee, and Kristina Toutanova, for coordinating with us on TACL presentations at ACL.
- Ani Nenkova and Owen Rambow, program co-chairs of NAACL 2016, and Michael Strube, program co-chair of ACL 2015, who were generous with advice.
- Yusuke Miyao, Yannick Versley, and Hai Zhao, our well-organized publication chairs, and the responsive team at Softconf led by Rich Gerber.
- Valia Kordoni and the local organization team, especially webmaster Kostadin Cholakov.
- Antal van den Bosch, our general chair, who kept us coordinated with the rest of the ACL 2016 team and offered guidance whenever we needed it.
- Antal van den Bosch, Claire Cardie, Pascale Fung, Ray Mooney, and Joakim Nivre, who carefully reviewed papers under consideration for outstanding and best paper recognition.
- Priscilla Rasmussen, who knows everything about how to make ACL a success.

We hope that you enjoy ACL 2016 in Berlin!

Katrin Erk, University of Texas
Noah A. Smith, University of Washington

Program Committee Co-Chairs

Organizing Committee

General Chair

Antal van den Bosch, Radboud University Nijmegen

Program Committee Co-chairs

Katrin Erk, University of Texas at Austin

Noah Smith, University of Washington

Local Organizing Committee

Valia Kordonis (local chair), Humboldt University

Markus Egg (local co-chair), Humboldt University

Kostadin Cholakov, Humboldt University

Maja Popović, Humboldt University

Manfred Stede, University of Potsdam

Anke Lüdeling, Humboldt University

Manfred Krifka, Humboldt University

Ulf Leser, Humboldt University

Workshop Co-chairs

Jun Zhao, Chinese Academy of Sciences

Sabine Schulte im Walde, IMS Stuttgart

Tutorial Co-chairs

Alexandra Birch, University of Edinburgh

Jelle Zuidema, University of Amsterdam

Publication Co-chairs

Yannick Versley, Ruprecht-Karls-Universität Heidelberg

Hai Zhao, Shanghai Jiao Tong University

Yusuke Miyao, National Institute of Informatics, Japan

Demo Co-chairs

Sameer Pradhan, Boulder Learning

Marianna Apidianaki, LIMSI-CNRS

Student Research Workshop Co-chairs

Will Roberts, Humboldt-Universität zu Berlin

Tao Lei, Massachusetts Institute of Technology

He He, University of Maryland

Faculty Advisors to the Student Research Workshop

Yang Liu, Tsinghua University

Chris Biemann, Technische Universität Darmstadt

Gosse Bouma, University of Groningen

Student Volunteer Coordinator

Yulia Grishina, University of Potsdam

Publicity Chair

Barbara Plank, University of Groningen

Business Manager

Priscilla Rasmussen

Conference Handbook Editors

Matt Post, Johns Hopkins University

Florian Kunneman, Radboud University Nijmegen

Professional Conference Organiser

Con gressa GmbH, Berlin

2

Tutorials: Sunday, August 7

Overview

7:30 – 6:00	Registration	(TBD)
7:30 – 9:00	Breakfast	(TBD)
9:00 – 12:30	Morning Tutorials	
	Multimodal Learning and Reasoning <i>Desmond Elliott, Douwe Kiela, and Angeliki Lazaridou</i>	(Governor's Square 15)
	NLP Approaches to Computational Argumentation <i>Noam Slonim, Iryna Girevycj, Chris Reed, and Benno Stein</i>	(Governor's Square 17)
	Computer Aided Translation <i>Philipp Koehn</i>	(Governor's Square 16)
	Semantic Representations of Word Senses and Concepts <i>José Camacho-Collados, Ignacio Iacobacci, Roberto Navigli, and Mohammad Taher Pilehvar</i>	(Governor's Square 17)
12:30 – 2:00	Lunch break	
2:00 – 5:30	Afternoon Tutorials	
	Neural Machine Translation <i>Thang Luong, Kyunghyun Cho, and Christopher D. Manning</i>	(Governor's Square 15)
	Game Theory and Natural Language: Origin, Evolution and Processing <i>Governor's Square 15)</i> <i>Rocco Tropodo and Marcello Pelillo</i>	
	Understanding Short Texts <i>Zhongyuan Wang and Haixun Wang</i>	(Governor's Square 15)

MetaNet: Repository, Identification System, and Applications
(Governor's Square 15)

Miriam R. L. Petrucci and Ellen K. Dodge

Tutorial 1

Multimodal Learning and Reasoning

Desmond Elliott, Douwe Kiela, and Angeliki Lazaridou

Sunday, August 7, 2016, 9:00–12:30pm

Governor's Square 15

Natural Language Processing has broadened in scope to tackle more and more challenging language understanding and reasoning tasks. The core NLP tasks remain predominantly unimodal, focusing on linguistic input, despite the fact that we, humans, acquire and use language while communicating in perceptually rich environments. Moving towards human-level AI will require the integration and modeling of multiple modalities beyond language. With this tutorial, our aim is to introduce researchers to the areas of NLP that have dealt with multimodal signals. The key advantage of using multimodal signals in NLP tasks is the complementarity of the data in different modalities. For example, we are less likely to find descriptions of yellow bananas or wooden chairs in text corpora, but these visual attributes can be readily extracted directly from images. Multimodal signals, such as visual, auditory or olfactory data, have proven useful for models of word similarity and relatedness, automatic image and video description, and even predicting the associated smells of words. Finally, multimodality offers a practical opportunity to study and apply multitask learning, a general machine learning paradigm that improves generalization performance of a task by using training signals of other related tasks.

Desmond Elliott is a postdoc at the Institute for Logic, Language and Computation in the University of Amsterdam (The Netherlands). His main research interests are models and evaluation methods for automatic image description. He delivered a tutorial on Datasets and Evaluation Methods for Image Description at the 2015 Integrating Vision and Language Summer School, and is co-organising a shared task on Multimodal Machine Translation at the 2016 Workshop on Machine Translation.

Douwe Kiela is a final year PhD student at the University of Cambridge's Computer Laboratory, supervised by Stephen Clark. He is interested in trying to enrich NLP with additional resources, primarily through grounding representations in perceptual modalities including vision, but also auditory and even olfactory modalities. He is a student board member of EACL and has published 8 top-tier conference papers over the three years of his PhD.

Angeliki Lazaridou is a final year PhD student, supervised by Marco Baroni at the Center for Mind/Brain Sciences of the University of Trento (Italy). Her primary research interests are in the area of multimodal semantics, i.e., making purely text-based models of meaning interact with other modalities, such as visual and sensorimotor. She has focused on learning models with multimodal signals and using those for multimodal inference, work that has appeared at related venues (ACL, NAACL, TACL, EMNLP).

Tutorial 2

NLP Approaches to Computational Argumentation

Noam Slonim, Iryna Girevycj, Chris Reed, and Benno Stein

Sunday, August 7, 2016, 9:00–12:30pm

Governor's Square 17

Argumentation and debating represent primary intellectual activities of the human mind. People in all societies argue and debate, not only to convince others of their own opinions but also in order to explore the differences between multiple perspectives and conceptualizations, and to learn from this exploration. The process of reaching a resolution on controversial topics typically does not follow a simple sequence of purely logical steps. Rather it involves a wide variety of complex and interwoven actions. Presumably, pros and cons are identified, considered, and weighed, via cognitive processes that often involve persuasion and emotions, which are inherently harder to formalize from a computational perspective.

This wide range of conceptual capabilities and activities, have only in part been studied in fields like CL and NLP, and typically within relatively small sub-communities that overlap the ACL audience. The new field of Computational Argumentation has very recently seen significant expansion within the CL and NLP community as new techniques and datasets start to become available, allowing for the first time investigation of the computational aspects of human argumentation in a holistic manner.

The main goal of this tutorial would be to introduce this rapidly evolving field to the CL community. Specifically, we will aim to review recent advances in the field and to outline the challenging research questions - that are most relevant to the ACL audience - that naturally arise when trying to model human argumentation.

We will further emphasize the practical value of this line of study, by considering real-world CL and NLP applications that are expected to emerge from this research, and to impact various industries, including legal, finance, healthcare, media, and education, to name just a few examples.

The first part of the tutorial will provide introduction to the basics of argumentation and rhetoric. Next, we will cover fundamental analysis tasks in Computational Argumentation, including argumentation mining, revealing argument relations, assessing arguments quality, stance classification, polarity analysis, and more. After the coffee break, we will first review existing resources and recently introduced benchmark data. In the following part we will cover basic synthesis tasks in Computational Argumentation, including the relation to NLG and dialogue systems, and the evolving area of Debate Technologies, defined as technologies developed directly to enhance, support, and engage with human debating. Finally, we will present relevant demos, review potential applications, and discuss the future of this emerging field.

Iryna Gurevych is Full Professor of Natural Language Processing in the Computer Science Department of the Technische Universitat (TU) Darmstadt. She heads the Ubiquitous Knowledge Processing (UKP) Lab and the Research Training Group ‘Adaptive Information Preparation from Heterogeneous Sources’ at the TU Darmstadt and Univ. of Heidelberg. Iryna has been working in the areas of lexical-semantic processing, discourse analysis and text mining with applications in social sciences and humanities for about 15 years. Her research interests have recently shifted towards NLP approaches to argumentation analysis. She is supervising several students working in this area and has published on this topic in the recent EMNLP, COLING and other conferences. Iryna has been co-organizer of the NAACL 2015 workshop on Argumentation Mining and the Dagstuhl seminar “Debating Technologies” in December 2015, she will co-organize the ACL 2016 workshop on Argument Mining in Berlin and co-chair the Debating Technologies track.

Chris Reed is Full Professor of Computer Science and Philosophy at the University of Dundee in Scotland, where he heads the Centre for Argument Technology. Chris has been working at the overlap between argumentation theory and artificial intelligence for over twenty years, has won over \$5.6m of funding from government and commercial sources and has over 140 peer-reviewed papers in the area including five books. He has also been instrumental in the development of the Argument Interchange Format, an international standard for computational work in the area; he is spear-heading the major engineering effort behind the Argument Web; and he is a founding editor of the Journal of Argument & Computation. He was co-organiser of COMMA 2014, of the first ACL workshop on Argumentation Mining in 2014, will be chair of the third workshop on Argument Mining with ACL in 2016, and has recently won funding for a \$1m project on the topic in collaboration with IBM.

Noam Slonim is a Senior Technical Staff Member (STSM) at IBM Research. He serves as the IBM Research Technical Lead of topics related to debate technologies, leading a team of more than 40 researchers from several different IBM Research Labs around the world. His main research interest is in developing innovative applications that can enhance, support, and engage with human debating. Correspondingly, he is actively pursuing specific research questions around that area, mainly in the context of developing advanced text analysis applications. From a theoretical perspective, his research over the years has led to the development of various Machine Learning techniques for the analysis of textual and genomic data, that often stem from Information Theoretic concepts and algorithms. Noam has recently published at ACL, EMNLP, and COLING.

Benno Stein is chair of the Web Technology and Information Systems Group at the Bauhaus-Universität Weimar. His research focuses on theory and algorithms for information retrieval, information extraction, and data mining. Since 2008 his group applies Big Data technology for research related to retrieval and extraction tasks. Among others, the group has developed effective and efficient methods for cross-language text reuse, document clustering, paraphrasing, and sentiment analysis. Benno is chair of the international workshop series TIR on Text-based Information Retrieval, initiator and co-organizer of PAN, a research network dedicated to digital text forensics. He is co-founder and spokesman of the Digital Bauhaus Lab, a forthcoming interdisciplinary research center with advanced working environments for research in Computer Science, Engineering, and Media Art and Design.

Tutorial 3

Computer Aided Translation

Philipp Koehn

Sunday, August 7, 2016, 9:00–12:30pm

Governor's Square 16

Moving beyond post-editing machine translation, a number of recent research efforts have advanced computer aided translation methods that allow for more interactivity, richer information such as confidence scores, and the completed feedback loop of instant adaptation of machine translation models to user translations.

This tutorial will explain the main techniques for several aspects of computer aided translation:

- confidence measures
- interactive machine translation (interactive translation prediction)
- bilingual concordancers
- translation option display
- paraphrasing (alternative translation suggestions)
- visualization of word alignment
- online adaptation
- automatic reviewing
- integration of translation memory
- eye tracking, logging, and cognitive user models

For each of these, the state of the art and open challenges are presented. The tutorial will also look under the hood of the open source CASMACAT toolkit that is based on MATECAT, and available as a “Home Edition” to be installed on a desktop machine. The target audience of this tutorials are researchers interested in computer aided machine translation and practitioners who want to use or deploy advanced CAT technology.

Philipp Koehn (John Hopkins University) is a leading researcher in statistical machine translation and co-ordinated the EU-funded CASMACAT project on computer aided translation 2011-2014.

Tutorial 4

Semantic Representations of Word Senses and Concepts

José Camacho-Collados, Ignacio Iacobacci, Roberto Navigli, and Mohammad Taher Pilehvar

Sunday, August 7, 2016, 9:00–12:30pm

Governor's Square 17

Representing the semantics of linguistic items in a machine interpretable form has been a major goal of Natural Language Processing since its earliest days. Among the range of different linguistic items, words have attracted the most research attention. However, word representations have an important limitation: they conflate different meanings of a word into a single vector. Representations of word senses have the potential to overcome this inherent limitation. Indeed, the representation of individual word senses and concepts has recently gained in popularity with several experimental results showing that a considerable performance improvement can be achieved across different NLP applications upon moving from word level to the deeper sense and concept levels. Another interesting point regarding the representation of concepts and word senses is that these models can be seamlessly applied to other linguistic items, such as words, phrases, sentences, etc.

This tutorial will first provide a brief overview of the recent literature concerning word representation (both count based and neural network based). It will then describe the advantages of moving from the word level to the deeper level of word senses and concepts, providing an extensive review of state of the art systems. Approaches covered will not only include those which draw upon knowledge resources such as WordNet, Wikipedia, BabelNet or FreeBase as reference, but also the so called multi prototype approaches which learn sense distinctions by using different clustering techniques. Our tutorial will discuss the advantages and potential limitations of all approaches, showing their most successful applications to date. We will conclude by presenting current open problems and lines of future work.

José Camacho Collados is a Google Doctoral Fellow and PhD student at the Sapienza University of Rome, working under the supervision of Prof. Roberto Navigli. His research focuses on Natural Language Processing and on the area of lexical semantics in particular. He has developed NASARI, a novel semantic vector representation for concepts and named entities, which led to two publications in NAACL and ACL 2015. José will co-organize a SemEval 2017 task on multilingual semantic similarity. His background education includes an Erasmus Mundus Master in Natural Language Processing and Human Language Technology and a 5-year BSc degree in Mathematics.

Ignacio Iacobacci is a PhD student at the Sapienza University of Rome, working under the supervision of Prof. Roberto Navigli. His research interests lie in the fields of Machine Learning, Natural Language Processing, Neural Networks. He is currently working on Word Sense Disambiguation and Distributional Semantics. Ignacio presented SensEmbed at ACL 2015, a novel approach for word and relational similarity built from exploiting semantic knowledge for modeling arbitrary word senses in a large sense inventory. His background includes a MSc. in Computer Science and 8 years as a developer including 4 years as a Machine Learning - NLP specialist.

Roberto Navigli is an Associate Professor in the Department of Computer Science at La Sapienza University of Rome and a member of the Linguistic Computing Laboratory. His research interests lie in the field of Natural Language Processing, including: Word Sense Disambiguation and Induction, Ontology Learning, Knowledge Representation and Acquisition, and multilinguality. In 2007 he received a Ph.D. in Computer Science from La Sapienza and he was awarded the Marco Cadoli 2007 AI*IA national prize for the Best Ph.D. Thesis in Artificial Intelligence. In 2013 he received the Marco Somalvico AI*IA prize, awarded every two years to the best young Italian researcher in Artificial Intelligence. He is the creator and founder of BabelNet, both a multilingual encyclopedic dictionary and a semantic network, and its related project Babelfy, a state-of-the-art multilingual disambiguation and entity linking system. He is also the Principal Investigator of MultiJEDI, a 1.3M euro 5-year Starting Grant funded by the European Research Council and the responsible person of the Sapienza unit in LIDER, an EU project on content analytics and language technologies. He is also the Co-PI of “Language Understanding cum Knowledge Yield” (LUcKY), a Google Focused Research Award on Natural Language Understanding.

Mohammad Taher Pilehvar is a Research Associate in the Language Technology Lab of the University of Cambridge where he is currently working on NLP in the biomedical domain. Taher completed his PhD in 2015 under the supervision of Prof. Roberto Navigli. Taher’s research lies in lexical semantics, mainly focusing on semantic representation, semantic similarity, and Word Sense Disambiguation. He has co-organized two semeval tasks and has authored multiple conference and journal papers on semantic representation and similarity in top tier venues. He is the first author of a paper on semantic similarity that was nominated for the best paper award at ACL 2013.

Tutorial 5

Neural Machine Translation

Thang Luong, Kyunghyun Cho, and Christopher D. Manning

Sunday, August 7, 2016, 2:00–5:30pm

Governor’s Square 15

Neural Machine Translation (NMT) is a simple new architecture for getting machines to learn to translate. Despite being relatively new (Kalchbrenner and Blunsom, 2013; Cho et al., 2014; Sutskever et al., 2014), NMT has already shown promising results, achieving state-of-the-art performances for various language pairs (Luong et al., 2015a; Jean et al., 2015; Luong et al., 2015b; Sennrich et al., 2016; Luong and Manning, 2016). While many of these NMT papers were presented to the ACL community, research and practice of NMT are only at their beginning stage. This tutorial would be a great opportunity for the whole community of machine translation and natural language processing to learn more about a very promising new approach to MT. This tutorial has four parts.

In the first part, we start with an overview of MT approaches, including: (a) traditional methods that have been dominant over the past twenty years and (b) recent hybrid models with the use of neural network components. From these, we motivate why an end-to-end approach like neural machine translation is needed. The second part introduces a basic instance of NMT. We start out with a discussion of recurrent neural networks, including the back-propagation-through-time algorithm and stochastic gradient descent optimizers, as these are the foundation on which NMT builds. We then describe in detail the basic sequence-to-sequence architecture of NMT (Cho et al., 2014; Sutskever et al., 2014), the maximum likelihood training approach, and a simple beam-search decoder to produce translations.

The third part of our tutorial describes techniques to build state-of-the-art NMT. We start with approaches to extend the vocabulary coverage of NMT (Luong et al., 2015a; Jean et al., 2015; Chitnis and DeNero, 2015). We then introduce the idea of jointly learning both translations and alignments through an attention mechanism (Bahdanau et al., 2015); other variants of attention (Luong et al., 2015b; Tu et al., 2016) are discussed too. We describe a recent trend in NMT, that is to translate at the sub-word level (Chung et al., 2016; Luong and Manning, 2016; Sennrich et al., 2016), so that language variations can be effectively handled. We then give tips on training and testing NMT systems such as batching and ensembling. In the final part of the tutorial, we briefly describe promising approaches, such as (a) how to combine multiple tasks to help translation (Dong et al., 2015; Luong et al., 2016; Firat et al., 2016; Zoph and Knight, 2016) and (b) how to utilize monolingual corpora (Sennrich et al., 2016). Lastly, we conclude with challenges remained to be solved for future NMT.

PS: we would also like to acknowledge the very first paper by Forcada and Neco (1997) on sequence-to-sequence models for translation!

Thang Luong is currently a 5th-year PhD student in the Stanford NLP group under Prof. Christopher Manning. In the past, he has published papers on various different NLP-related areas such as digital library, machine translation, speech recognition, parsing, psycholinguistics, and word embedding learning. Recently, his main interest shifts towards the area of deep learning using sequence to sequence models to tackle various NLP problems, especially neural machine translation. He has built state-of-the-art (academically) neural machine translation systems both at Google and at Stanford.

Kyunghyun Cho is an assistant professor in the Department of Computer Science and the Center for Data Science at New York University. He has worked in deep learning for natural language processing, language translation, image captioning, and a variety of other subjects including the core methods of deep learning. He completed a post-doctoral fellowship at the Montreal Institute for Learning Algorithms. He earned a Ph.D and an M.Sc (with distinction) from the Aalto University School of Science, and a B.Sc in Computer Science from the Korea Advanced Institute of Science and Technology.

Christopher Manning is a professor of computer science and linguistics at Stanford University. He works on software that can intelligently process, understand, and generate human language material. He is a leader in applying Deep Learning to Natural Language Processing, including exploring Tree Recursive Neural Networks, sentiment analysis, neural network dependency parsing, the GloVe model of word vectors, neural machine translation, and deep language understanding. Manning is an ACM Fellow, a AAAI Fellow, and an ACL Fellow, and has coauthored leading textbooks on statistical natural language processing and information retrieval. He is a member of the Stanford NLP group (@stanfordnlp).

Tutorial 6

Game Theory and Natural Language: Origin, Evolution and Processing

Rocco Tropodo and Marcello Pelillo

Sunday, August 7, 2016, 2:00–5:30pm

Governor's Square 16

The development of game theory in the early 1940's by John von Neumann was a reaction against the then dominant view that problems in economic theory can be formulated using standard methods from optimization theory. Indeed, most real-world economic problems involve conflicting interactions among decision-making agents that cannot be adequately captured by a single (global) objective function. The main idea behind game theory is to shift the emphasis from optimality criteria to equilibrium conditions. Game theory provides a framework to model complex scenarios, with applications in economics and social science but also in different fields of information technology. With the recent development of algorithmic game theory, it has been used to solve problems in computer vision, pattern recognition, machine learning and natural language processing.

Game-theoretic frameworks have been used in different ways to study language origin and evolution. Furthermore, the so-called game metaphor has been used by philosophers and linguists to explain how language evolved and how it works. Ludwig Wittgenstein, for example, famously introduced the concept of a language game to explain the conventional nature of language, and put forward the idea of the spontaneous formation of a common language that gradually emerges from the interactions among the speakers within a population.

This concept opens the way to the interpretation of language as a complex adaptive system composed of linguistic units and their interactions, which gives rise to the emergence of structural properties. It is the core part of many computational models of language that are based on classical game theory and evolutionary game theory. With the former it is possible to model how speakers form a signaling system in which the ambiguity of the symbols is minimized; with the latter it is possible to model how speakers coordinate their linguistic choices according to the satisfaction that they have about the outcome of a communication act, converging to a common language. In the same vein, many other attempts have been proposed to explain how other characteristics of language follow similar dynamics.

Game theory, and in particular evolutionary game theory, thanks to their ability to model interactive situations and to integrate information from multiple sources, have also been used to solve specific problems in natural language processing and information retrieval, such as language generation, word sense disambiguation and document and text clustering.

The goal of this tutorial is to offer an introduction to the basic concepts of game theory and to show its main applications in the study of language, from different perspectives. We shall assume no pre-existing knowledge of game theory by the audience, thereby making the tutorial self-contained and understandable by a non-expert.

Marcello Pelillo is Full Professor of Computer Science at Ca' Foscari University, where he directs the European Centre for Living Technology (ECLT). He held visiting research positions at Yale University, McGill University, the University of Vienna, York University (UK), the University College London, and the National ICT Australia (NICTA). He has published more than 200 technical papers in refereed journals, handbooks, and conference proceedings in the areas of machine learning, pattern recognition and computer vision. He has initiated several conference series, including EMMCVPR in 1997, IWCV in 2008, SIMBAD in 2011, and he chairs the EMMCVPR and SIMBAD steering committees. He serves (has served) on the Editorial Boards of the journals IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), IET Computer Vision, Pattern Recognition, Brain Informatics, and he serves on the Advisory Board of the International Journal of Machine Learning and Cybernetics. He has served (serves) as Guest Editor for various special issues of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Pattern Recognition, Pattern Recognition Letters. He is (or has been) scientific coordinator of several research projects, including SIMBAD, an EU-FP7 project devoted to similarity-based pattern analysis and recognition whose activity is described in a recently published Springer book, and he has recently won an award from the Samsung Global Research Outreach (GRO) program. Prof. Pelillo is a Fellow of the IEEE and a Fellow of the IAPR. He has recently been appointed IEEE SMC Distinguished Lecturer.

Rocco Tripodi is Post-doctoral researcher at European Centre for Living Technology (Ca' Foscari University of Venice), where he is working on learning models based on game theoretic principles. He completed his PhD in Computer Science at Ca' Foscari University in 2015 with a thesis titled “Evolutionary Game Theoretic Models for Natural Language Processing”. He was Research Assistant and Adjunct Professor at Ca' Foscari University in 2010, where he worked on ontological semantics and taught Corpus Linguistics and Natural Language Processing for Online Applications. In 2015, he was visiting researcher at Staffordshire University (UK). His research interests are in the areas of machine learning and natural language processing. He is particularly interested in classification problems on networked data, from a semantic perspective and on the design, learning and evolution of linguistic communication systems. At present, he is working on game theoretic learning models, with applications to the analysis of multi-modal signals from interacting agents and on group dynamics.

Tutorial 7

Understanding Short Texts

Zhongyuan Wang and Haixun Wang

Sunday, August 7, 2016, 2:00–5:30pm

Billions of short texts are produced every day, in the form of search queries, ad keywords, tags, tweets, messenger conversations, social network posts, etc. Unlike documents, short texts have some unique characteristics which make them difficult to handle. First, short texts, especially search queries, do not always observe the syntax of a written language. This means traditional NLP techniques, such as syntactic parsing, do not always apply to short texts. Second, short texts contain limited context. The majority of search queries contain less than 5 words, and tweets can have no more than 140 characters. Because of the above reasons, short texts give rise to a significant amount of ambiguity, which makes them extremely difficult to handle. On the other hand, many applications, including search engines, ads, automatic question answering, online advertising, recommendation systems, etc., rely on short text understanding. In all these applications, the necessary first step is to transform an input text into a machine-interpretable representation, namely to “understand” the short text. A growing number of approaches leverage external knowledge to address the issue of inadequate contextual information that accompanies the short texts. These approaches can be classified into two categories: Explicit Representation Model (ERM) and Implicit Representation Model (IRM). In this tutorial, we will present a comprehensive overview of short text understanding based on explicit semantics (knowledge graph representation, acquisition, and reasoning) and implicit semantics (embedding and deep learning). Specifically, we will go over various techniques in knowledge acquisition, representation, and inferencing has been proposed for text understanding, and we will describe massive structured and semi-structured data that have been made available in the recent decade that directly or indirectly encode human knowledge, turning the knowledge representation problems into a computational grand challenge with feasible solutions insight.

Zhongyuan Wang is a Researcher at Microsoft Research Asia (MSRA). He leads two projects at MSRA: Enterprise Dictionary (knowledge mining from Enterprise) and Probbase (knowledge mining from Web). He got his Ph.D. degree in computer science from Renmin University of China, and his PhD thesis is ‘Short Text Understanding’. Zhongyuan Wang has published 20+ papers (including ICDE 2015 Best Paper Award on short text understanding) in the leading international conferences, such as VLDB, ICDE, IJCAI, CIKM, etc. He is also the co-author of the book ‘Web Data Management: Concepts and Techniques’, published in 2014. His research interests include knowledge base, natural language processing, semantic network, machine learning, and web data mining.

Haixun Wang is a research scientist / Engineering manager at Facebook. Before Facebook, he was with Google Research, working on natural language processing. He led research in semantic search, graph data processing systems, and distributed query processing at Microsoft Research Asia. He had been a research staff member at IBM T. J. Watson Research Center from 2000 - 2009. He was Technical Assistant to Stuart Feldman (Vice President of Computer Science of IBM Research) from 2006 to 2007, and Technical Assistant to Mark Wegman (Head of Computer Science of IBM Research) from 2007 to 2009. He received the Ph.D. degree in computer science from the University of California, Los Angeles in 2000. He has published more than 150 research papers in refereed international journals and conference proceedings. He served as PC Chair of conferences such as CIKM12 and he is on the editorial board of IEEE Transactions of Knowledge and Data Engineering (TKDE), and Journal of Computer Science and Technology (JCST). He won the best paper award in ICDE 2015, 10 year best paper award in ICDM 2013, and best paper award of ER 2009.

Tutorial 8

MetaNet: Repository, Identification System, and Applications

Miriam R. L. Petrucci and Ellen K. Dodge

Sunday, August 7, 2016, 2:00–5:30pm

The ubiquity of metaphor in language (Lakoff and Johnson 1980) has served as impetus for cognitive linguistic approaches to the study of language, mind, and the study of mind (e.g. Thibodeau & Boroditsky 2011). While native speakers use metaphor naturally and easily, the treatment and interpretation of metaphor in computational systems remains challenging because such systems have not succeeded in developing ways to recognize the semantic elements that define metaphor. This tutorial demonstrates MetaNet’s frame-based semantic analyses, and their informing of MetaNet’s automatic metaphor identification system. Participants will gain a complete understanding of the theoretical basis and the practical workings of MetaNet, and acquire relevant information about the Frame Semantics basis of that knowledge base and the way that FrameNet handles the widespread phenomenon of metaphor in language. The tutorial is geared to researchers and practitioners of language technology, not necessarily experts in metaphor analysis or knowledgeable about either FrameNet or MetaNet, but who are interested in natural language processing tasks that involve automatic metaphor processing, or could benefit from exposure to tools and resources that support frame-based deep semantic, analyses of language, including metaphor as a widespread phenomenon in human language.

Miriam R. L. Petrucci received her PhD in Linguistics from the University of California, Berkeley. A key member of the team developing FrameNet almost since the project’s founding, her research interests include semantics, knowledge base development, grammar and lexis, lexical semantics, Frame Semantics and Construction Grammar.

Ellen K. Dodge received her PhD in Linguistics from the University of California, Berkeley. Since 2000, she has worked in ICSI’s AI Group, first as part of the Neural Theory of Language project, and also FrameNet. She is the primary linguist continuing to develop Embodied Construction Grammar. Since 2012, she has worked on MetaNet, developing formal representations of frame and metaphor networks, as well as automatic methods to identify and analyze metaphors in text.

Main Conference: Monday, June 1

Overview

8:45–9:00	Opening session							
9:00–10:10	Invited talk I: Amber Boydston					<i>Plaza Ballroom A, B, & C</i>		
10:10–10:40	Coffee break							
10:40–12:00	Session 1	Semantic parsing I	Information extraction	Machine translation I	Word meaning I	Parsing I	Noncompositionality	
12:00–1:40	Lunch break							
1:40–3:00	Session 2	Word vectors I	Events and schemas	Sentiment analysis	Parsing II	Information retrieval	Phonology and morphology	
3:00–3:30	Coffee break							
3:30–5:10	Session 3	Question answering I	Sentence vectors	Parsing III	Dialog	Generation	Entities and coreference	Topics
6:00–9:00	Poster and dinner session I (includes SRW)						<i>Plaza Ballroom A, B, & C</i>	
6:00–9:00	ACL System Demonstrations Session A							

Session 1 Overview – Monday, June 1, 2015

	Track A <i>Semantic parsing I</i>	Track B <i>Information extraction</i>	Track C <i>Machine translation I</i>	Track D <i>Word meaning I</i>	Track E <i>Parsing I</i>	Track F <i>Noncompositionality</i>
	Plaza Ballroom	#1 Ballroom	#1 Ballroom	F		
10:40	Noise reduction and targeted exploration in imitation learning for Abstract Meaning Representation parsing <i>Goodman, Vlachos, and Naradowsky</i>	Unsupervised Person Slot Filling based on Graph Mining <i>Yu and Ji</i>	Models and Inference for Prefix-Constrained Machine Translation <i>Wuebker, Green, DeNero, Hasan, and Luong</i>	Incremental Acquisition of Verb Hypothesis Space towards Physical World Interaction <i>She and Chai</i>	Transition-based dependency parsing with topological fields <i>Kok and Hinrichs</i>	Literal and Metaphorical Senses in Compositional Distributional Semantic Models <i>Gutierrez, Shutova, Marghetis, and Bergen</i>
11:00	Data Recombination for Neural Semantic Parsing <i>Jia and Liang</i>	Online Adaptor Grammars with Hybrid inference <i>Delli Bovi, Telesca, and Navigli</i>	Modeling Coverage for Neural Machine Translation <i>Tu, Lu, Liu, Liu, and Li</i>	Language Transfer Learning for Supervised Lexical Substitution <i>Hintz and Biemann</i>	Generalized Transition-based Dependency Parsing via Control Parameters <i>Bohnet, McDonald, Pitler, and Ma</i>	Idiom Token Classification using Sentential Distributed Semantics <i>Salton, Ross, and Kelleher</i>
11:20	Inferring Logical Forms From Denotations <i>Pasupat and Liang</i>	A Multi-media Approach to Cross-lingual Entity Knowledge Transfer <i>Lu, Pan, Pourdamghani, Chang, Ji, and Knight</i>	Improving Neural Machine Translation Models with Monolingual Data <i>Sennrich, Haddow, and Birch</i>	Learning the Curriculum with Bayesian Optimization for Task-Specific Word Representation Learning <i>Tsvetkov, Faruqui, Ling, MacWhinney, and Dyer</i>	A Transition-Based System for Joint Lexical and Syntactic Analysis <i>Constant and Nivre</i>	Adaptive Joint Learning of Compositional and Non-Compositional Phrase Embeddings <i>Hashimoto and Tsuruoka</i>
11:40	Language to Logical Form with Neural Attention <i>Dong and Lapata</i>	Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations <i>Chiu and Nichols</i>	Graph-Based Translation Via Graph Segmentation <i>Li, Way, and Liu</i>	Pointing the Unknown Words <i>Gulcehre, Ahn, Nallapati, Zhou, and Bengio</i>	Neural Greedy Constituent Parsing with Dynamic Oracles <i>Coavoux and Crabbé</i>	Metaphor Detection with Topic Transition, Emotion and Cognition in Context <i>Jang, Jo, Shen, Miller, Moon, and Rose</i>

Parallel Session 1

1A: Semantic parsing I

Plaza Ballroom A & B

Chair:

Noise reduction and targeted exploration in imitation learning for Abstract Meaning Representation parsing

James Goodman, Andreas Vlachos, and Jason Naradowsky

10:40–11:00

Semantic parsers map natural language statements into meaning representations, and must abstract over syntactic phenomena, resolve anaphora, and identify word senses to eliminate ambiguous interpretations. Abstract meaning representation (AMR) is a recent example of one such semantic formalism which, similar to a dependency parse, utilizes a graph to represent relationships between concepts (Banarescu et al., 2013). As with dependency parsing, transition-based approaches are a common approach to this problem. However, when trained in the traditional manner these systems are susceptible to the accumulation of errors when they find undesirable states during greedy decoding. Imitation learning algorithms have been shown to help these systems recover from such errors. To effectively use these methods for AMR parsing we find it highly beneficial to introduce two novel extensions: noise reduction and targeted exploration. The former mitigates the noise in the feature representation, a result of the complexity of the task. The latter targets the exploration steps of imitation learning towards areas which are likely to provide the most information in the context of a large action-space. We achieve state-of-the-art results, and improve upon standard transition-based parsing by 4.7 F1 points.

Data Recombination for Neural Semantic Parsing

Robin Jia and Percy Liang

11:00–11:20

Modeling crisp logical regularities is crucial in semantic parsing, making it difficult for neural models with no task-specific prior knowledge to achieve good results. In this paper, we introduce data recombination, a novel framework for injecting such prior knowledge into a model. From the training data, we induce a high-precision synchronous context-free grammar, which captures important conditional independence properties commonly found in semantic parsing. We then train a sequence-to-sequence recurrent network (RNN) model with a novel attention-based copying mechanism on datapoints sampled from this grammar, thereby teaching the model about these structural properties. Data recombination improves the accuracy of our RNN model on three semantic parsing datasets, leading to new state-of-the-art performance on the standard GeoQuery dataset for models with comparable supervision.

Inferring Logical Forms From Denotations

Panupong Pasupat and Percy Liang

11:20–11:40

A core problem in learning semantic parsers from denotations is picking out consistent logical forms—those that yield the correct denotation—from a combinatorially large space. To control the search space, previous work relied on restricted set of rules, which limits expressivity. In this paper, we consider a much more expressive class of logical forms, and show how to use dynamic programming to efficiently represent the complete set of consistent logical forms. Expressivity also introduces many more spurious logical forms which are consistent with the correct denotation but do not represent the meaning of the utterance. To address this, we generate fictitious worlds and use crowdsourced denotations on these worlds to filter out spurious logical forms. On the WikiTableQuestions dataset, we increase the coverage of answerable questions from 53.5% to 76%, and the additional crowdsourced supervision lets us rule out 92.1% of spurious logical forms.

Language to Logical Form with Neural Attention

Li Dong and Mirella Lapata

11:40–12:00

Semantic parsing aims at mapping natural language to machine interpretable meaning representations. Traditional approaches rely on high-quality lexicons, manually-built templates, and linguistic features which are either domain- or representation-specific. In this paper we present a general method based on an attention-enhanced encoder-decoder model. We encode input utterances into vector representations, and generate their logical forms by conditioning the output sequences or trees on the encoding vectors. Experimental results on four datasets show that our approach performs competitively without using hand-engineered features and is easy to adapt across domains and meaning representations.

1B: Information extraction

Plaza Ballroom D & E

*Chair:***Unsupervised Person Slot Filling based on Graph Mining***Dian Yu and Heng Ji*

10:40–11:00

Slot filling aims to extract the values (slot fillers) of specific attributes (slots types) for a given entity (query) from a large-scale corpus. Slot filling remains very challenging over the past seven years. We propose a simple yet effective unsupervised approach to extract slot fillers based on the following two observations: (1) a trigger is usually a salient node relative to the query and filler nodes in the dependency graph of a context sentence; (2) a relation is likely to exist if the query and candidate filler nodes are strongly connected by a relation-specific trigger. Thus we design a graph-based algorithm to automatically identify triggers based on personalized PageRank and Affinity Propagation for a given (query, filler) pair and then label the slot type based on the identified triggers. Our approach achieves 11.6%–25% higher F-score over state-of-the-art English slot filling methods. Our experiments also demonstrate that as long as a few trigger seeds, name tagging and dependency parsing capabilities exist, this approach can be quickly adapted to any language and new slot types. Our promising results on Chinese slot filling can serve as a new benchmark.

Online Adaptor Grammars with Hybrid inference*Claudio Delli Bovi, Luca Telesca, and Roberto Navigli*

11:00–11:20

We present DefIE, an approach to large-scale Information Extraction (IE) based on a syntactic-semantic analysis of textual definitions. Given a large corpus of definitions we leverage syntactic dependencies to reduce data sparsity, then disambiguate the arguments and content words of the relation strings, and finally exploit the resulting information to organize the acquired relations hierarchically. The output of DefIE is a high-quality knowledge base consisting of several million automatically acquired semantic relations.

A Multi-media Approach to Cross-lingual Entity Knowledge Transfer*Di Lu, Xiaoman Pan, Nima Pourdamghani, Shih-Fu Chang, Heng Ji, and Kevin Knight* 11:20–11:40

When a large-scale incident or disaster occurs, there is often a great demand for rapidly developing a system to extract detailed and new information from low-resource languages (LLs). We propose a novel approach to discover comparable documents in high-resource languages (HLs), and project Entity Discovery and Linking results from HLs documents back to LLs. We leverage a wide variety of language-independent forms from multiple data modalities, including image processing (image-to-image retrieval, visual similarity and face recognition) and sound matching. We also propose novel methods to learn entity priors from a large-scale HL corpus and knowledge base. Using Hausa and Chinese as the LLs and English as the HL, experiments show that our approach achieves 36.1% higher Hausa name tagging F-score over a costly supervised model, and 9.4% higher Chinese-to-English Entity Linking accuracy over state-of-the-art.

Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations*Jason P.C. Chiu and Eric Nichols*

11:40–12:00

Named entity recognition is a challenging task that has traditionally required large amounts of knowledge in the form of feature engineering and lexicons to achieve high performance. In this paper, we present a novel neural network architecture that automatically detects word- and character-level features using a hybrid bidirectional LSTM and CNN architecture, eliminating the need for most feature engineering. We also propose a novel method of encoding partial lexicon matches in neural networks and compare it to existing approaches. Extensive evaluation shows that, given only tokenized text and publicly available word embeddings, our system is competitive on the CoNLL-2003 dataset and surpasses the previously reported state of the art on the OntoNotes 5.0 dataset by 2.13 F1 points. By using two lexicons from public sources, we establish new states of the art with an F1 score of 91.62 on CoNLL-2003 and 86.28 on OntoNotes, surpassing systems that employ heavy feature engineering, proprietary lexicons, and rich entity linking information.

1C: Machine translation I

Plaza Ballroom F

Chair:

Models and Inference for Prefix-Constrained Machine Translation*Joern Wuebker, Spence Green, John DeNero, Sasa Hasan, and Minh-Thang Luong* 10:40–11:00

We apply phrase-based and neural models to a core task in interactive machine translation: suggesting how to complete a partial translation. For the phrase-based system, we demonstrate improvements in suggestion quality using novel objective functions, learning techniques, and inference algorithms tailored to this task. Our contributions include new tunable metrics, an improved beam search strategy, an n-best extraction method that increases suggestion diversity, and a tuning procedure for a hierarchical joint model of alignment and translation. The combination of these techniques improves next-word suggestion accuracy dramatically from 28.5% to 41.2% in a large-scale English-German experiment. Our recurrent neural translation system increases accuracy to 53.0%, but inference is two orders of magnitude slower. Manual error analysis shows the strengths and weaknesses of both approaches.

Modeling Coverage for Neural Machine Translation*Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li*

11:00–11:20

Attention mechanism has enhanced state-of-the-art Neural Machine Translation (NMT) by jointly learning to align and translate. It tends to ignore past alignment information, however, which often leads to over-translation and under-translation. To address this problem, we propose coverage-based NMT in this paper. We maintain a coverage vector to keep track of the attention history. The coverage vector is fed to the attention model to help adjust future attention, which lets NMT system to consider more about untranslated source words. Experiments show that the proposed approach significantly improves both translation quality and alignment quality over standard attention-based NMT.

Improving Neural Machine Translation Models with Monolingual Data*Rico Sennrich, Barry Haddow, and Alexandra Birch*

11:20–11:40

Neural Machine Translation (NMT) has obtained state-of-the-art performance for several language pairs, while only using parallel data for training. Target-side monolingual data plays an important role in boosting fluency for phrase-based statistical machine translation, and we investigate the use of monolingual data for NMT. In contrast to previous work, which combines NMT models with separately trained language models, we note that encoder-decoder NMT architectures already have the capacity to learn the same information as a language model, and we explore strategies to train with monolingual data without changing the neural network architecture. By pairing monolingual training data with an automatic back-translation, we can treat it as additional parallel training data, and we obtain substantial improvements on the WMT 15 task English->German (+2.8–3.7 BLEU), and for the low-resourced IWSLT 14 task Turkish->English (+2.1–3.4 BLEU), obtaining new state-of-the-art results. We also show that fine-tuning on in-domain monolingual and parallel data gives substantial improvements for the IWSLT 15 task English-German.

Graph-Based Translation Via Graph Segmentation*Liangyou Li, Andy Way, and Qun Liu*

11:40–12:00

One major drawback of phrase-based translation is that it segments an input sentence into continuous phrases. To support linguistically informed source discontinuity, in this paper we construct graphs which combine bigram and dependency relations and propose a graph-based translation model. The model segments an input graph into connected subgraphs, each of which may cover a discontinuous phrase. We use beam search to combine translations of each subgraph left-to-right to produce a complete translation. Experiments on Chinese–English and German–English tasks show that our system is significantly better than the phrase-based model by up to +1.5/+0.5 BLEU scores. By explicitly modeling the graph segmentation, our system obtains further improvement, especially on German–English.

1D: Word meaning I

Chair:

Incremental Acquisition of Verb Hypothesis Space towards Physical World Interaction

Lanbo She and Joyce Chai

10:40–11:00

As a new generation of cognitive robots start to enter our lives, it is important to enable robots to follow human commands and to learn new actions from human language instructions. To address this issue, this paper presents an approach that explicitly represents verb semantics through hypothesis spaces of fluents and automatically acquires these hypothesis spaces by interacting with humans. The learned hypothesis spaces can be used to automatically plan for lower-level primitive actions towards physical world interaction. Our empirical results have shown that the representation of a hypothesis space of fluents, combined with the learned hypothesis selection algorithm, outperforms a previous baseline. In addition, our approach applies incremental learning which can contribute to life-long learning from humans in the future.

Language Transfer Learning for Supervised Lexical Substitution

Gerold Hintz and Chris Biemann

11:00–11:20

We propose a framework for lexical substitution that is able to perform transfer learning across languages. Datasets for this task are available in at least three languages (English, Italian, and German). Previous work has addressed each of these tasks in isolation. In contrast, we regard the union of three shared tasks as a combined multilingual dataset. We show that a supervised system can be trained effectively, even if training and evaluation data are from different languages. Successful transfer learning between languages suggests that the learned model is in fact independent of the underlying language. We combine state-of-the-art unsupervised features obtained from syntactic word embeddings and distributional thesauri in a supervised delexicalized ranking system. Our system improves over state of the art in the full lexical substitution task in all three languages.

Learning the Curriculum with Bayesian Optimization for Task-Specific Word Representation Learning

Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Brian MacWhinney, and Chris Dyer 11:20–11:40

We use Bayesian optimization to learn curricula for word representation learning, optimizing performance on downstream tasks that depend on the learned representations as features. The curricula are modeled by a linear ranking function which is the scalar product of a learned weight vector and an engineered feature vector that characterizes the different aspects of the complexity of each instance in the training corpus. We show that learning the curriculum improves performance on a variety of downstream tasks over random orders and in comparison to the natural corpus order.

Pointing the Unknown Words

Caglar Gulcehre, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio 11:40–12:00

The problem of rare and unknown words is an important issue that can potentially effect the performance of many NLP systems, including traditional count-based and deep learning models. We propose a novel way to deal with the rare and unseen words for the neural network models using attention. Our model uses two softmax layers in order to predict the next word in conditional language models: one predicts the location of a word in the source sentence, and the other predicts a word in the shortlist vocabulary. At each timestep, the decision of which softmax layer to use is adaptively made by an MLP which is conditioned on the context.~We motivate this work from a psychological evidence that humans naturally have a tendency to point towards objects in the context or the environment when the name of an object is not known.~Using our proposed model, we observe improvements on two tasks, neural machine translation on the Europarl English to French parallel corpora and text summarization on the Gigaword dataset.

1E: Parsing I

Chair:

Transition-based dependency parsing with topological fields

Daniel de Kok and Erhard Hinrichs

10:40–11:00

The topological field model is commonly used to describe the regularities in German word order. In this work, we show that topological fields can be predicted reliably using sequence labeling and that the predicted field labels can improve transition-based dependency parsing.

Generalized Transition-based Dependency Parsing via Control Parameters

Bernd Bohnet, Ryan McDonald, Emily Pitler, and Ji Ma

11:00–11:20

In this paper, we present a generalized transition-based parsing framework where parsers are instantiated in terms of a set of control parameters that constrain transitions between parser states. This generalization provides a unified framework to describe and compare various transition-based parsing approaches from both a theoretical and empirical perspective. This includes well-known transition-systems, but also previously un-studied systems.

A Transition-Based System for Joint Lexical and Syntactic Analysis

Matthieu Constant and Joakim Nivre

11:20–11:40

We present a transition-based system that jointly predicts the syntactic structure and lexical units of a sentence by building two structures over the input words: a syntactic dependency tree and a forest of lexical units including multiword expressions (MWEs). This combined representation allows us to capture both the syntactic and semantic structure of MWEs, which in turn enables deeper downstream semantic analysis, especially for semi-compositional MWEs. The proposed system extends the arc-standard transition system for dependency parsing with transitions for building complex lexical units. Experiments on two different data sets show that the approach significantly improves MWE identification accuracy (and sometimes syntactic accuracy) compared to existing joint approaches.

Neural Greedy Constituent Parsing with Dynamic Oracles

Maximin Coavoux and Benoit Crabbe

11:40–12:00

Dynamic oracle training has shown substantial improvements for dependency parsing in various settings, but has not been explored for constituent parsing. The present article introduces a dynamic oracle for transition-based constituent parsing. Experiments on the 9 languages of the SPMRL dataset show that a neural greedy parser with morphological features, trained with a dynamic oracle, leads to accuracies comparable with the best non-reranking and non-ensemble parsers.

1F: Noncompositionality

Chair:

Literal and Metaphorical Senses in Compositional Distributional Semantic Models

E.Dario Gutierrez, Ekaterina Shutova, Tyler Marghetis, and Benjamin Bergen 10:40–11:00

Metaphorical expressions are pervasive in natural language and pose a substantial challenge for computational semantics. The inherent compositionality of metaphor makes it an important test case for compositional distributional semantic models (CDSMs). This paper is the first to investigate whether metaphorical composition warrants a distinct treatment in the CDSM framework. We propose a method to learn metaphors as linear transformations in a vector space and find that, across a variety of semantic domains, explicitly modeling metaphor improves the resulting semantic representations. We then use these representations in a metaphor identification task, achieving a high performance of 0.82 in terms of F-score.

Idiom Token Classification using Sentential Distributed Semantics

Giancarlo Salton, Robert Ross, and John Kelleher 11:00–11:20

Idiom token classification is the task of deciding for a set of potentially idiomatic phrases whether each occurrence of a phrase is a literal or idiomatic usage of the phrase. In this work we explore the use of Skip-Thought Vectors to create distributed representations that encode features that are predictive with respect to idiom token classification. We show that classifiers using these representations have competitive performance compared with the state of the art in idiom token classification. Importantly, however, our models use only the sentence containing the target phrase as input and are thus less dependent on a potentially inaccurate or incomplete model of discourse context. We further demonstrate the feasibility of using these representations to train a competitive general idiom token classifier.

Adaptive Joint Learning of Compositional and Non-Compositional Phrase Embeddings

Kazuma Hashimoto and Yoshimasa Tsuruoka 11:20–11:40

We present a novel method for jointly learning compositional and non-compositional phrase embeddings by adaptively weighting both types of embeddings using a compositionality scoring function. The scoring function is used to quantify the level of compositionality of each phrase, and the parameters of the function are jointly optimized with the objective for learning phrase embeddings. In experiments, we apply the adaptive joint learning method to the task of learning embeddings of transitive verb phrases, and show that the compositionality scores have strong correlation with human ratings for verb-object compositionality, substantially outperforming the previous state of the art. Moreover, our embeddings improve upon the previous best model on a transitive verb disambiguation task. We also show that a simple ensemble technique further improves the results for both tasks.

Metaphor Detection with Topic Transition, Emotion and Cognition in Context

Hyeju Jang, Yohan Jo, Qinlan Shen, Michael Miller, Seungwhan Moon, and Carolyn Rose 11:40–12:00

Metaphor is a common linguistic tool in communication, making its detection in discourse a crucial task for natural language understanding. One popular approach to this challenge is to capture semantic incohesion between a metaphor and the dominant topic of the surrounding text. While these methods are effective, they tend to overclassify target words as metaphorical when they deviate in meaning from its context. We present a new approach that (1) distinguishes literal and non-literal use of target words by examining sentence-level topic transitions and (2) captures the motivation of speakers to express emotions and abstract concepts metaphorically. Experiments on an online breast cancer discussion forum dataset demonstrate a significant improvement in metaphor detection over the state-of-the-art. These experimental results also reveal a tendency toward metaphor usage in personal topics and certain emotional contexts.

Session 2 Overview – Monday, June 1, 2015

Track A Word vectors I	Track B Events and schemas	Track C Sentiment analysis	Track D Parsing II	Track E Information retrieval	Track F Phonology and morphology
Plaza Ballroom	#1Ballroom	#1Ballroom	F		
Word Embeddings as Metric Recovery in Semantic Spaces <i>Hashimoto, Alvarez-Melis, and Jaakkola</i>	Liberal Event Extraction and Event Schema Induction <i>Huang, Cassidy, Feng, Ji, Voss, Han, and Sil</i>	Sentiment Domain Adaptation with Multiple Sources <i>Wu and Huang</i>	Active Learning for Dependency Parsing with Partial Annotation <i>Li, Zhang, Zhang, Liu, Chen, Wu, and Wang</i>	Query Expansion with Locally-Trained Word Embeddings <i>Diaz, Mitra, and Craswell</i>	Grapheme-to-Phoneme Models for (Almost) Any Language <i>Deri and Knight</i>
Compressing Neural Language Models by Sparse Word Representations <i>Chen, Mou, Xu, Li, and Jin</i>	Jointly Event Extraction and Visualization on Twitter via Probabilistic Modelling <i>ZHOU, Gao, and He</i>	Connotation Frames: A Data-Driven Investigation <i>Rashkin, Singh, and Choi</i>	Online Adaptor Grammars with Hybrid inference <i>Kuhlmann and Jonsson</i>	Together we stand: Siamese Networks for Similar Question Retrieval <i>Das, Yenala, Chinnakotla, and Shrivastava</i>	Online Adaptor Grammars with Hybrid inference <i>Faruqui, McDonald, and Soricev</i>
Intrinsic Subspace Evaluation of Word Embedding Representations <i>Yaghoobzadeh and Schütze</i>	Using Sentence-Level LSTM Language Models for Script Inference <i>Pichotta and Mooney</i>	Bi-Transferring Deep Neural Networks for Domain Adaptation <i>Zhou, Xie, Huang, and He</i>	Dependency Parsing with Bounded Block Degree and Well-nestedness via Lagrangian Relaxation and Branch-and-Bound <i>Corro, Le Roux, Lacroix, Rozenknap, and Wolfner Calvo</i>	News Citation Recommendation with Implicit and Explicit Semantics <i>Peng, Liu, and Lin</i>	Neural Word Segmentation Learning for Chinese <i>Cai and Zhao</i>
On the Role of Seed Lexicons in Learning Bilingual Word Embeddings <i>Vulić and Korhonen</i>	Two Discourse-Driven Language Models for Semantics <i>Peng and Roth</i>	Document-level Sentiment Inference with Social, Faction, and Discourse Context <i>Choi, Rashkin, Zettlemoyer, and Choi</i>		Scalable Semi-Supervised Query Classification Using Matrix Sketching <i>Kim, Stratos, and Sarikaya</i>	Transition-Based Neural Word Segmentation <i>Zhang, Zhang, and Fu</i>

Parallel Session 2

2A: Word vectors I

Plaza Ballroom A & B

Chair:

Word Embeddings as Metric Recovery in Semantic Spaces

Tatsunori Hashimoto, David Alvarez-Melis, and Tommi Jaakkola

13:40–14:00

Continuous word representations have been remarkably useful across NLP tasks but remain poorly understood. We ground word embeddings in semantic spaces studied in the cognitive-psychometric literature, taking these spaces as the primary objects to recover. To this end, we relate log co-occurrences of words in large corpora to semantic similarity assessments and show that co-occurrences are indeed consistent with an Euclidean semantic space hypothesis. Framing word embedding as metric recovery of a semantic space unifies existing word embedding algorithms, ties them to manifold learning, and demonstrates that existing algorithms are consistent metric recovery methods given co-occurrence counts from random walks. Furthermore, we propose a simple, principled, direct metric recovery algorithm that performs on par with the state-of-the-art word embedding and manifold learning methods. Finally, we complement recent focus on analogies by constructing two new inductive reasoning datasets—series completion and classification—and demonstrate that word embeddings can be used to solve them as well.

Compressing Neural Language Models by Sparse Word Representations

Yunchuan Chen, Lili Mou, Yan Xu, Ge Li, and Zhi Jin

14:00–14:20

Neural networks are among the state-of-the-art techniques for language modeling. Existing neural language models typically map discrete words to distributed, dense vector representations. After information processing of the preceding context words by hidden layers, an output layer estimates the probability of the next word. Such approaches are time- and memory-intensive because of the large numbers of parameters for word embeddings and the output layer. In this paper, we propose to compress neural language models by sparse word representations. In the experiments, the number of parameters in our model increases very slowly with the growth of the vocabulary size, which is almost imperceptible. Moreover, our approach not only reduces the parameter space to a large extent, but also improves the performance in terms of the perplexity measure.

Intrinsic Subspace Evaluation of Word Embedding Representations

Yadollah Yaghoobzadeh and Hinrich Schütze

14:20–14:40

We introduce a new methodology for intrinsic evaluation of word representations. Specifically, we identify four fundamental criteria based on the characteristics of natural language that pose difficulties to NLP systems; and develop tests that directly show whether or not representations contain the subspaces necessary to satisfy these criteria. Current intrinsic evaluations are mostly based on the overall similarity or full-space similarity of words and thus view vector representations as points. We show the limits of these point-based intrinsic evaluations. We apply our evaluation methodology to the comparison of a count vector model and several neural network models and demonstrate important properties of these models.

On the Role of Seed Lexicons in Learning Bilingual Word Embeddings

Ivan Vulic and Anna Korhonen

14:40–15:00

A shared bilingual word embedding space (SBWES) is an indispensable resource in a variety of cross-language NLP and IR tasks. A common approach to the SBWES induction is to learn a mapping function between monolingual semantic spaces, where the mapping critically relies on a seed word lexicon used in the learning process. In this work, we analyze the importance and properties of seed lexicons for the SBWES induction across different dimensions (i.e., lexicon source, lexicon size, translation method, translation pair reliability). On the basis of our analysis, we propose a simple but effective hybrid bilingual word embedding (BWE) model. This model (HYBWE) learns the mapping between two monolingual embedding spaces using only highly reliable symmetric translation pairs from a seed document-level embedding space. We perform bilingual lexicon learning (BLL) with 3 language pairs and show that by carefully selecting reliable translation pairs our new HYBWE model outperforms benchmarking BWE learning models, all of which use more expensive bilingual signals. Effectively, we demonstrate that a SBWES may be induced by leveraging only a very weak bilingual signal (document alignments) along with monolingual data.

2B: Events and schemas

Plaza Ballroom D & E

Chair:

Liberal Event Extraction and Event Schema Induction

Lifu Huang, Taylor Cassidy, Xiaocheng Feng, Heng Ji, Clare R. Voss, Jiawei Han, and Avirup Sil 13:40–14:00

We propose a brand new “Liberal” Event Extraction paradigm to extract events and discover event schemas from any input corpus simultaneously. We incorporate symbolic (e.g., Abstract Meaning Representation) and distributional semantics to detect and represent event structures and adopt a joint typing framework to simultaneously extract event types and argument roles and discover an event schema. Experiments on general and specific domains demonstrate that this framework can construct high-quality schemas with many event and argument role types, covering a high proportion of event types and argument roles in manually defined schemas. We show that extraction performance using discovered schemas is comparable to supervised models trained from a large amount of data labeled according to pre-defined event types. The extraction quality of new event types is also promising.

Jointly Event Extraction and Visualization on Twitter via Probabilistic Modelling

Deyu ZHOU, Tianmeng Gao, and Yulan He

14:00–14:20

Event extraction from texts aims to detect structured information such as what has happened, to whom, where and when. Event extraction and visualization are typically considered as two different tasks. In this paper, we propose a novel approach based on probabilistic modelling to jointly extract and visualize events from tweets where both tasks benefit from each other. We model each event as a joint distribution over named entities, a date, a location and event-related keywords. Moreover, both tweets and event instances are associated with coordinates in the visualization space. The manifold assumption that the intrinsic geometry of tweets is a low-rank, non-linear manifold within the high-dimensional space is incorporated into the learning framework using a regularization. Experimental results show that the proposed approach can effectively deal with both event extraction and visualization and performs remarkably better than both the state-of-the-art event extraction method and a pipeline approach for event extraction and visualization.

Using Sentence-Level LSTM Language Models for Script Inference

Karl Pichotta and Raymond J. Mooney

14:20–14:40

There is a small but growing body of research on statistical scripts, models of event sequences that allow probabilistic inference of implicit events from documents. These systems operate on structured verb-argument events produced by an NLP pipeline. We compare these systems with recent Recurrent Neural Net models that directly operate on raw tokens to predict sentences, finding the latter to be roughly comparable to the former in terms of predicting missing events in documents.

Two Discourse Driven Language Models for Semantics

Haoruo Peng and Dan Roth

14:40–15:00

Natural language understanding often requires deep semantic knowledge. Expanding on previous proposals, we suggest that some important aspects of semantic knowledge can be modeled as a language model if done at an appropriate level of abstraction. We develop two distinct models that capture semantic frame chains and discourse information while abstracting over the specific mentions of predicates and entities. For each model, we investigate four implementations: a “standard” N-gram language model and three discriminatively trained “neural” language models that generate embeddings for semantic frames. The quality of the semantic language models (SemLM) is evaluated both intrinsically, using perplexity and a narrative cloze test and extrinsically – we show that our SemLM helps improve performance on semantic natural language processing tasks such as co-reference resolution and discourse parsing.

2C: Sentiment analysis

Plaza Ballroom F

Chair:

Sentiment Domain Adaptation with Multiple Sources*Fangzhao Wu and Yongfeng Huang*

13:40–14:00

Domain adaptation is an important research topic in sentiment analysis area. Existing domain adaptation methods usually transfer sentiment knowledge from only one source domain to target domain. In this paper, we propose a new domain adaptation approach which can exploit sentiment knowledge from multiple source domains. We first extract both global and domain-specific sentiment knowledge from the data of multiple source domains using multi-task learning. Then we transfer them to target domain with the help of words' sentiment polarity relations extracted from the unlabeled target domain data. The similarities between target domain and different source domains are also incorporated into the adaptation process. Experimental results on benchmark dataset show the effectiveness of our approach in improving cross-domain sentiment classification performance.

Connotation Frames: A Data-Driven Investigation*Hannah Rashkin, Sameer Singh, and Yejin Choi*

14:00–14:20

Through a particular choice of a predicate (e.g., “x violated y”), a writer can subtly connote a range of implied sentiment and presupposed facts about the entities x and y: (1) writer’s perspective: projecting x as an “antagonist” and y as a “victim”, (2) entities’ perspective: y probably dislikes x, (3) effect: something bad happened to y, (4) value: y is something valuable, and (5) mental state: y is distressed by the event. We introduce connotation frames as a representation formalism to organize these rich dimensions of connotation using typed relations. First, we investigate the feasibility of obtaining connotative labels through crowdsourcing experiments. We then present models for predicting the connotation frames of verb predicates based on their distributional word representations and the interplay between different types of connotative relations. Empirical results confirm that connotation frames can be induced from various data sources that reflect how language is used in context. We conclude with analytical results that show the potential use of connotation frames for analyzing subtle biases in online news media.

Bi-Transferring Deep Neural Networks for Domain Adaptation*Guangyou Zhou, Zhiwen Xie, Jimmy Xiangji Huang, and Tingting He*

14:20–14:40

Sentiment classification aims to automatically predict sentiment polarity (e.g., positive or negative) of user generated sentiment data (e.g., reviews, blogs). Due to the mismatch among different domains, a sentiment classifier trained in one domain may not work well when directly applied to other domains. Thus, domain adaptation for sentiment classification algorithms are highly desirable to reduce the domain discrepancy and manual labeling costs. To address the above challenge, we propose a novel domain adaptation method, called Bi-Transferring Deep Neural Networks (BTDNNs). The proposed BTDNNs attempts to transfer the source domain examples to the target domain, and also transfer the target domain examples to the source domain. The linear transformation of BTDNNs ensures the feasibility of transferring between domains, and the distribution consistency between the transferred domain and the desirable domain is constrained with a linear data reconstruction manner. As a result, the transferred source domain is supervised and follows similar distribution as the target domain. Therefore, any supervised method can be used on the transferred source domain to train a classifier for sentiment classification in a target domain. We conduct experiments on a benchmark composed of reviews of 4 types of Amazon products. Experimental results show that our proposed approach significantly outperforms the several baseline methods, and achieves an accuracy which is competitive with the state-of-the-art method for domain adaptation.

Document-level Sentiment Inference with Social, Faction, and Discourse Context*Eunsol Choi, Hannah Rashkin, Luke Zettlemoyer, and Yejin Choi*

14:40–15:00

We present a new approach for document-level sentiment inference, where the goal is to predict directed opinions (who feels positively or negatively towards whom) for all entities mentioned in a text. To encourage more complete and consistent predictions, we introduce an ILP that jointly models (1) sentence- and discourse-level sentiment cues, (2) factual evidence about entity factions, and (3) global constraints based on social science theories such as homophily, social balance, and reciprocity. Together, these cues allow for rich inference across groups of entities, including for example that CEOs and the companies they lead are likely to have similar sentiment towards others. We evaluate performance on new, densely labeled data that provides supervision for all pairs, complementing previous work that only labeled pairs mentioned in the same sentence. Experi-

iments demonstrate that the global model outperforms sentence-level baselines, by providing more coherent predictions across sets of related entities.

2D: Parsing II*Chair:***Active Learning for Dependency Parsing with Partial Annotation***Zhenghua Li, Min Zhang, Yue Zhang, Zhanyi Liu, Wenliang Chen, Hua Wu, and Haifeng Wang*
13:40–14:00

Different from traditional active learning based on sentence-wise full annotation (FA), this paper proposes active learning with dependency-wise partial annotation (PA) as a finer-grained unit for dependency parsing. At each iteration, we select a few most uncertain words from an unlabeled data pool, manually annotate their syntactic heads, and add the partial trees into labeled data for parser retraining. Compared with sentence-wise FA, dependency-wise PA gives us more flexibility in task selection and avoids wasting time on annotating trivial tasks in a sentence. Our work makes the following contributions. First, we are the first to apply a probabilistic model to active learning for dependency parsing, which can 1) provide tree probabilities and dependency marginal probabilities as principled uncertainty metrics, and 2) directly learn parameters from PA based on a forest-based training objective. Second, we propose and compare several uncertainty metrics through simulation experiments on both Chinese and English. Finally, we conduct human annotation experiments to compare FA and PA on real annotation time and quality.

Online Adaptor Grammars with Hybrid inference*Marco Kuhlmann and Peter Jonsson*

14:00–14:20

We study the generalization of maximum spanning tree dependency parsing to maximum acyclic subgraphs. Because the underlying optimization problem is intractable even under an arc-factored model, we consider the restriction to noncrossing dependency graphs. Our main contribution is a cubic-time exact inference algorithm for this class. We extend this algorithm into a practical parser and evaluate its performance on four linguistic data sets used in semantic dependency parsing. We also explore a generalization of our parsing framework to dependency graphs with pagenumber at most $\$k\$$ and show that the resulting optimization problem is NP-hard for $\$k \geq 2\$$.

Dependency Parsing with Bounded Block Degree and Well-nestedness via Lagrangian Relaxation and Branch-and-Bound*Caio Corro, Joseph Le Roux, Mathieu Lacroix, Antoine Rozenknop, and Roberto Wolfert Calvo*
14:20–14:40

We present a novel dependency parsing method which enforces two structural properties on dependency trees: bounded block degree and well-nestedness. These properties are useful to better represent the set of admissible dependency structures in treebanks and connect dependency parsing to context-sensitive grammatical formalisms. We cast this problem as an Integer Linear Program that we solve with Lagrangian Relaxation from which we derive a heuristic and an exact method based on a Branch-and-Bound search. Experimentally, we see that these methods are efficient and competitive compared to a baseline unconstrained parser, while enforcing structural properties in all cases.

2E: Information retrieval*Chair:***Query Expansion with Locally-Trained Word Embeddings***Fernando Diaz, Bhaskar Mitra, and Nick Craswell*

13:40–14:00

Continuous space word embeddings have received a great deal of attention in the natural language processing and machine learning communities for their ability to model term similarity and other relationships. We study the use of term relatedness in the context of query expansion for ad hoc information retrieval. We demonstrate that word embeddings such as word2vec and GloVe, when trained globally, underperform corpus and query specific embeddings for retrieval tasks. These results suggest that other tasks benefiting from global embeddings may also benefit from local embeddings.

Together we stand: Siamese Networks for Similar Question Retrieval*Arpita Das, Harish Yenala, Manoj Chinnakotla, and Manish Shrivastava*

14:00–14:20

Community Question Answering (cQA) services like Yahoo! Answers, Baidu Zhidao, Quora, StackOverflow etc. provide a platform for interaction with experts and help users to obtain precise and accurate answers to their questions. The time lag between the user posting a question and receiving its answer could be reduced by retrieving similar historic questions from the cQA archives. The main challenge in this task is the “lexico-syntactic” gap between the current and the previous questions. In this paper, we propose a novel approach called “Siamese Convolutional Neural Network for cQA (SCQA)” to find the semantic similarity between the current and the archived questions. SCQA consist of twin convolutional neural networks with shared parameters and a contrastive loss function joining them. SCQA learns the similarity metric for question-question pairs by leveraging the question-answer pairs available in cQA forum archives. The model projects semantically similar question pairs nearer to each other and dissimilar question pairs farther away from each other in the semantic space. Experiments on large scale real-life “Yahoo! Answers” dataset reveals that SCQA outperforms current state-of-the-art approaches based on translation models, topic models and deep neural network based models which use non-shared parameters.

News Citation Recommendation with Implicit and Explicit Semantics*Hao Peng, Jing Liu, and Chin-Yew Lin*

14:20–14:40

In this work, we focus on the problem of news citation recommendation. The task aims to recommend news citations for both authors and readers to create and search news references. Due to the sparsity issue of news citations and the engineering difficulty in obtaining information on authors, we focus on content similarity-based methods instead of collaborative filtering-based approaches. In this paper, we explore word embedding (i.e. implicit semantics) and grounded entities (i.e. explicit semantics) to address the variety and ambiguity issues of language. We formulate the problem as a re-ranking task and integrate different similarity measures under the learning to rank framework. We evaluate our approach on a real-world dataset. The experimental results show the efficacy of our method.

Scalable Semi-Supervised Query Classification Using Matrix Sketching*Young-Bum Kim, Karl Stratos, and Ruhi Sarikaya*

14:40–15:00

The enormous scale of unlabeled text available today necessitates scalable schemes for representation learning in natural language processing. For instance, in this paper we are interested in classifying the intent of a user query. While our labeled data is quite limited, we have access to virtually an unlimited amount of unlabeled queries, which could be used to induce useful representations: for instance by principal component analysis(PCA). However, it is prohibitive to even store the data in memory due to its sheer size, let alone apply conventional batch algorithms. In this work, we apply the recently proposed matrix sketching algorithm to entirely obviate the problem with scalability (Liberty, 2013). This algorithm approximates the data within a specified memory bound while preserving the covariance structure necessary for PCA. Using matrix sketching, we significantly improve the user intent classification accuracy by leveraging large amounts of unlabeled queries.

2F: Phonology and morphology*Chair:***Grapheme-to-Phoneme Models for (Almost) Any Language***Alyya Deri and Kevin Knight*

13:40–14:00

Grapheme-to-phoneme (g2p) models are rarely available in low-resource languages, as the creation of training and evaluation data is expensive and time-consuming. We use Wiktionary to obtain more than 650k word-pronunciation pairs in more than 500 languages. We then develop phoneme and language distance metrics based on phonological and linguistic knowledge; applying those, we adapt g2p models for high-resource languages to create models for related low-resource languages. We provide results for models for 229 adapted languages.

Online Adaptor Grammars with Hybrid inference*Manaal Faruqui, Ryan McDonald, and Radu Soricut*

14:00–14:20

Morpho-syntactic lexicons provide information about the morphological and syntactic roles of words in a language. Such lexicons are not available for all languages and even when available, their coverage can be limited. We present a graph-based semi-supervised learning method that uses the morphological, syntactic and semantic relations between words to automatically construct wide coverage lexicons from small seed sets. Our method is language-independent, and we show that we can expand a 1000 word seed lexicon to more than 100 times its size with high quality for 11 languages. In addition, the automatically created lexicons provide features that improve performance in two downstream tasks: morphological tagging and dependency parsing.

Neural Word Segmentation Learning for Chinese*Deng Cai and Hai Zhao*

14:20–14:40

Most previous approaches to Chinese word segmentation formalize this problem as a character-based sequence labeling task so that only contextual information within fixed sized local windows and simple interactions between adjacent tags can be captured. In this paper, we propose a novel neural framework which thoroughly eliminates context windows and can utilize complete segmentation history. Our model employs a gated combination neural network over characters to produce distributed representations of word candidates, which are then given to a long short-term memory (LSTM) language scoring model. Experiments on the benchmark datasets show that without the help of feature engineering as most existing approaches, our models achieve competitive or better performances with previous state-of-the-art methods.

Transition-Based Neural Word Segmentation*Meishan Zhang, Yue Zhang, and Guohong Fu*

14:40–15:00

Character-based and word-based methods are two main types of statistical models for Chinese word segmentation, the former exploiting sequence labeling models over characters and the latter typically exploiting a transition-based model, with the advantages that word-level features can be easily utilized. Neural models have been exploited for character-based Chinese word segmentation, giving high accuracies by making use of external character embeddings, yet requiring less feature engineering. In this paper, we study a neural model for word-based Chinese word segmentation, by replacing the manually-designed discrete features with neural features in a word-based segmentation framework. Experimental results demonstrate that word features lead to comparable performances to the best systems in the literature, and a further combination of discrete and neural features gives top accuracies.

Session 3 Overview – Monday, June 1, 2015

	Track A Question answering I <i>Plaza Ballroom</i>	Track B Sentence vectors <i>R&B Ballroom</i>	Track C Parsing III <i>D&E Ballroom</i>	Track D Dialog <i>F</i>	Track E Generation	Track F Entities and coreference	Track G Topics
03:30	A Parallel-Hierarchical Model for Machine Comprehension on Sparse Data <i>Trischler, Ye, Yuan, He, and Bachman</i>	Neural Summarization by Extracting Sentences and Words <i>Cheng and Lapata</i>	Word Embeddings as Metric Recovery in Semantic Spaces <i>Agić, Johannsen, Plank, Martínez Alonso, Schlüter, and Søgaard</i>	DocChat: An Information Retrieval Approach for Chatbot Engines Using Unstructured Documents <i>Yan, Duan, Bao, Chen, Zhou, Li, and Zhou</i>	Towards more variation in text generation: Developing and evaluating variation models for choice of referential form <i>Castro Ferreira, Kraemer, and Wubben</i>	Collective Entity Resolution with Multi-Focal Attention <i>Globerson, Lazic, Chakrabarti, Subramanya, Ringgaard, and Pereira</i>	Effects of Creativity and Cluster Tightness on Short Text Clustering Performance <i>Finegan-Dollak, Coke, Zhang, Ye, and Radev</i>
03:50	Combining Natural Logic and Shallow Reasoning for Question Answering <i>Angeli, Nayak, and Manning</i>	Word Embeddings as Metric Recovery in Semantic Spaces <i>Yin, Schütze, Xiang, and Zhou</i>	Online Adaptor Grammars with Hybrid inference <i>Zhai, Boyd-Graber, and Cohen</i>	Investigating the Sources of Linguistic Alignment in Conversation <i>Doyle and Frank</i>	How Much is 131 Million Dollars? Putting Numbers in Perspective with Compositional Descriptions <i>Chaganty and Liang</i>	Online Adaptor Grammars with Hybrid inference <i>Lazic, Subramanya, Ringgaard, and Pereira</i>	Generative Topic Embedding: a Continuous Representation of Documents <i>Li, Chua, Zhu, and Miao</i>
04:10	Easy Questions First? A Case Study on Curriculum Learning for Question Answering <i>Sachan and Xing</i>	Neural Networks For Negation Scope Detection <i>Fancellu, Lopez, and Webber</i>	Many Languages, One Parser <i>Ammar, Mulcaire, Ballesteros, Dyer, and Smith</i>	Entropy Converges Between Dialogue Participants: Explanations from an Information-Theoretic Perspective <i>Xu and Reitter</i>	Generating Factoid Questions With Recurrent Neural Networks: The 30M Factoid Question-Answer Corpus <i>Serban, García-Durán, Gulcehre, Ahn, Chandar, Courville, and Bengio</i>	Online Adaptor Grammars with Hybrid inference <i>Nguyen, Theobald, and Weikum</i>	Detecting Cross-cultural Differences Using a Multilingual Topic Model <i>Gutierrez, Shutova, Lichtenstein, Melo, and Gilardi</i>
04:30	Improved Representation Learning for Question Answer Matching <i>Tan, Santos, Xiang, and Zhou</i>	CSE: Conceptual Sentence Embeddings based on Attention Model <i>Huang, Feng, Zhou, Gu, and Gao</i>	Word Embeddings as Metric Recovery in Semantic Spaces <i>Stratos, Collins, and Hsu</i>	Finding the Middle Ground - A Model for Planning Satisficing Answers <i>Janzen, Maqî, and Kowatsch</i>	Latent Predictor Networks for Code Generation <i>Ling, Blunsom, Grefenstette, Hermann, Kočiský, Wang, and Senior</i>	Which Coreference Evaluation Metric Do You Trust? A Proposal for a Link-based Entity Aware Metric <i>Moosavi and Strube</i>	Detecting Common Discussion Topics Across Culture From News Reader Comments <i>Shi, Lam, Bing, and Xu</i>

04:50	Tables as Semi-structured Knowledge for Question Answering <i>Jauhar, Turney, and Hovy</i>	Word Embeddings as Metric Recovery in Semantic Spaces <i>Gulordava and Merlo</i>	A Sentence Interaction Network for Modeling Dependence between Sentences <i>Liu and Huang</i>	Easy Things First: Installments Improve Referring Expression Generation for Objects in Photographs <i>Zarrieß and Schlangen</i>	Improving Coreference Resolution by Learning Entity-Level Distributed Representations <i>Clark and Manning</i>	A Discriminative Topic Model using Document Network Structure <i>Yang, Boyd-Graber, and Resnik</i>
-------	---	---	--	--	---	---

Parallel Session 3

3A: Question answering I

Plaza Ballroom A & B

Chair:

A Parallel-Hierarchical Model for Machine Comprehension on Sparse Data

Adam Trischler, Zheng Ye, Xingdi Yuan, Jing He, and Philip Bachman

15:30–15:50

Understanding unstructured text is a major goal within natural language processing. Comprehension tests pose questions based on short text passages to evaluate such understanding. In this work, we investigate machine comprehension on the challenging MCTest benchmark. Partly because of its limited size, prior work on MCTest has focused mainly on engineering better features. We tackle the dataset with a neural approach, harnessing simple neural networks arranged in a parallel hierarchy. The parallel hierarchy enables our model to compare the passage, question, and answer from a variety of trainable perspectives, as opposed to using a manually designed, rigid feature set. Perspectives range from the word level to sentence fragments to sequences of sentences; the networks operate only on word-embedding representations of text. When trained with a methodology designed to help cope with limited training data, our Parallel-Hierarchical model sets a new state of the art for MCTest, outperforming previous feature-engineered approaches slightly and previous neural approaches by a significant margin (over 15 percentage points)

Combining Natural Logic and Shallow Reasoning for Question Answering

Gabor Angeli, Neha Nayak, and Christopher D. Manning

15:50–16:10

Broad-domain question answering often requires shallow textual methods in order to gain coverage beyond facts that are available in structured knowledge bases, but simultaneously can benefit from the high precision of logical reasoning. We propose an approach for incorporating both of these signals in a unified framework based on natural logic. We train an evaluation function — akin to gameplaying — to evaluate the expected truth of candidate premises on the fly. As an added contribution, we extend the breadth of inferences afforded by natural logic to include relational entailment (e.g., $\text{buy} \rightarrow \text{own}$) and meronymy (e.g., a person born in a city is born in the city's country). We evaluate our approach on answering multiple-choice science questions, outperforming prior work on the dataset.

Easy Questions First? A Case Study on Curriculum Learning for Question Answering

Mrinmaya Sachan and Eric Xing

16:10–16:30

Cognitive science researchers have emphasized the importance of ordering a complex task into a sequence of easy to hard problems. Such an ordering provides an easier path to learning and increases the speed of acquisition of the task compared to conventional learning. Recent works in machine learning have explored a curriculum learning approach called self-paced learning which orders data samples on the easiness scale so that easy samples can be introduced to the learning algorithm first and harder samples can be introduced successively. We explore curriculum learning in the context of non-convex models for question answering and show that these approaches lead to improvements. We argue that incorporating easy, yet, a diverse set of questions, which are different from the questions already seen by the learner, can further improve learning. Our experiments support this hypothesis.

Improved Representation Learning for Question Answer Matching

Ming Tan, Cicero dos Santos, Bing Xiang, and Bowen Zhou

16:30–16:50

Passage-level question answer matching is a challenging task since it requires effective representations that capture the complex semantic relations between questions and answers. In this work, we propose a series of deep learning models to address passage answer selection. To match passage answers to questions accommodating their complex semantic relations, unlike most previous work that utilizes a single deep learning structure, we develop hybrid models that process the text using both convolutional and recurrent neural networks, combining the merits on extracting linguistic information from both structures. Additionally, we also develop a simple but effective attention mechanism for the purpose of constructing better answer representations according to the input question, which is imperative for better modeling long answer sequences. The results on two public benchmark datasets, InsuranceQA and TREC-QA, show that our proposed models outperform a variety of strong baselines.

Tables as Semi-structured Knowledge for Question Answering

Sujay Kumar Jauhar, Peter Turney, and Eduard Hovy

16:50–17:10

Question answering requires access to a knowledge base to check facts and reason about information. Knowledge in the form of natural language text is easy to acquire, but difficult for automated reasoning. Highly-structured knowledge bases can facilitate reasoning, but are difficult to acquire. In this paper we explore tables as a semi-structured formalism that provides a balanced compromise to this trade-off. We first use the structure of tables to guide the construction of a dataset of over 9000 multiple-choice questions with rich alignment annotations, easily and efficiently via crowd-sourcing. We then use this annotated data to train a semi-structured feature-driven model for question answering that uses tables as a knowledge base. In benchmark evaluations, we significantly outperform both a strong unstructured retrieval baseline and a highly-structured Markov Logic Network model.

3B: Sentence vectors

Plaza Ballroom D & E

Chair:

Neural Summarization by Extracting Sentences and Words*Jianpeng Cheng and Mirella Lapata*

15:30–15:50

Traditional approaches to extractive summarization rely heavily on human-engineered features. In this work we propose a data-driven approach based on neural networks and continuous sentence features. We develop a general framework for single-document summarization composed of a hierarchical document encoder and an attention-based extractor. This architecture allows us to develop different classes of summarization models which can extract sentences or words. We train our models on large scale corpora containing hundreds of thousands of document-summary pairs. Experimental results on two summarization datasets demonstrate that our models obtain results comparable to the state of the art without any access to linguistic annotation.

Word Embeddings as Metric Recovery in Semantic Spaces*Wenpeng Yin, Hinrich Schütze, Bing Xiang, and Bowen Zhou*

15:50–16:10

How to model a pair of sentences is a critical issue in many NLP tasks such as answer selection (AS), paraphrase identification (PI) and textual entailment (TE). Most prior work (i) deals with one individual task by fine-tuning a specific system; (ii) models each sentence's representation separately, rarely considering the impact of the other sentence; or (iii) relies fully on manually designed, task-specific linguistic features. This work presents a general Attention Based Convolutional Neural Network (ABCNN) for modeling a pair of sentences. We make three contributions. (i) The ABCNN can be applied to a wide variety of tasks that require modeling of sentence pairs. (ii) We propose three attention schemes that integrate mutual influence between sentences into CNNs; thus, the representation of each sentence takes into consideration its counterpart. These interdependent sentence pair representations are more powerful than isolated sentence representations. (iii) ABCNNs achieve state-of-the-art performance on AS, PI and TE tasks.

Neural Networks For Negation Scope Detection*Federico Fancellu, Adam Lopez, and Bonnie Webber*

16:10–16:30

Automatic negation scope detection is a task that has been tackled using different classifiers and heuristics. Most systems are however 1) highly-engineered, 2) English-specific, and 3) only tested on the same genre they were trained on. We start by addressing 1) and 2) using a neural network architecture. Results obtained on data from the *SEM2012 shared task on negation scope detection show that even a simple feed-forward neural network using word-embedding features alone, performs on par with earlier classifiers, with a bi-directional LSTM outperforming all of them. We then address 3) by means of a specially-designed synthetic test set; in doing so, we explore the problem of detecting the negation scope more in depth and show that performance suffers from genre effects and differs with the type of negation considered.

CSE: Conceptual Sentence Embeddings based on Attention Model*Heyan Huang, Chong Feng, Qiang Zhou, Jiahui Gu, and Xiong Gao*

16:30–16:50

Most sentence embedding models typically represent each sentence only using word surface, which makes these models indiscriminative for ubiquitous polysemy. In order to enhance semantic representation capability of sentence, we employ text conceptualization algorithm to assign associated concepts for each sentence, and then learn conceptual sentence embedding (CSE). Therefore, this concept-level representation is more expressive than some widely-used text representation models such as latent topic model, especially for short-text. Moreover, we further extend CSE models by utilizing a local attention-based model that select relevant words within the context to make more efficient prediction. The experimental results on text classification task and information retrieval task, show that the proposed models outperform the topic model-based sentence representation methods and other sentence embedding models.

3C: Parsing III

Plaza Ballroom F

*Chair:***Word Embeddings as Metric Recovery in Semantic Spaces***Željko Agić, Anders Johannsen, Barbara Plank, Héctor Martínez Alonso, Natalie Schluter, and Anders Søgaard*

15:30–15:50

We propose a novel approach to cross-lingual part-of-speech tagging and dependency parsing for truly low-resource languages. Our annotation projection-based approach yields tagging and parsing models for over 100 languages. All that is needed are freely available parallel texts, and taggers and parsers for resource-rich languages. The empirical evaluation across 30 test languages shows that our method consistently provides top-level accuracies, close to established upper bounds, and outperforms several competitive baselines.

Online Adaptor Grammars with Hybrid inference*Ke Zhai, Jordan Boyd-Graber, and Shay B. Cohen*

15:50–16:10

Adaptor grammars are a flexible, powerful formalism for defining nonparametric, unsupervised models of grammar productions. This flexibility comes at the cost of expensive inference. We address the difficulty of inference through an online algorithm which uses a hybrid of Markov chain Monte Carlo and variational inference. We show that this inference strategy improves scalability without sacrificing performance on unsupervised word segmentation and topic modeling tasks.

Many Languages, One Parser*Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah A. Smith* 16:10–16:30

We train one multilingual model for dependency parsing and use it to parse sentences in several languages. The parsing model uses (i) multilingual word clusters and embeddings; (ii) token-level language information; and (iii) language-specific features (fine-grained POS tags). This input representation enables the parser not only to parse effectively in multiple languages, but also to generalize across languages based on linguistic universals and typological similarities, making it more effective to learn from limited annotations. Our parser's performance compares favorably to strong baselines in a range of data scenarios, including when the target language has a large treebank, a small treebank, or no treebank for training.

Word Embeddings as Metric Recovery in Semantic Spaces*Karl Stratos, Michael Collins, and Daniel Hsu*

16:30–16:50

We tackle unsupervised part-of-speech (POS) tagging by learning hidden Markov models (HMMs) that are particularly well-suited for the problem. These HMMs, which we call anchor HMMs, assume that each tag is associated with at least one word that can have no other tag, which is a relatively benign condition for POS tagging (e.g., “the” is a word that appears only under the determiner tag). We exploit this assumption and extend the non-negative matrix factorization framework of Arora et al. (2013) to design a consistent estimator for anchor HMMs. In experiments, our algorithm is competitive with strong baselines such as the clustering method of Brown et al. (1992) and the log-linear model of Berg-Kirkpatrick et al. (2010). Furthermore, it produces an interpretable model in which hidden states are automatically lexicalized by words.

Word Embeddings as Metric Recovery in Semantic Spaces*Kristina Gulordava and Paola Merlo*

16:50–17:10

The growing work in multi-lingual parsing faces the challenge of fair comparative evaluation and performance analysis across languages and their treebanks. The difficulty lies in teasing apart the properties of treebanks, such as their size or average sentence length, from those of the annotation scheme, and from the linguistic properties of languages. We propose a method to evaluate the effects of word order of language on dependency parsing performance, while controlling for confounding treebank properties. The method uses artificially-generated treebanks that are minimal permutations of actual treebanks with respect to two word order properties: word order variation and dependency lengths. Based on these artificial data on twelve languages, we show that longer dependencies and higher word order variability degrade parsing performance. Our method also extends to minimal pairs of individual sentences, leading to a finer-grained understanding of parsing errors.

3D: Dialog*Chair:***DocChat: An Information Retrieval Approach for Chatbot Engines Using Unstructured Documents***Zhao Yan, Nan Duan, Junwei Bao, Peng Chen, Ming Zhou, Zhoujun Li, and Jianshe Zhou* 15:30–15:50

Most current chatbot engines are designed to reply to user utterances based on existing utterance-response (or U-R) pairs. In this paper, we present DocChat, a novel information retrieval approach for chatbot engines that can leverage unstructured documents, instead of U-R pairs, to respond to utterances. A learning to rank model with features designed at different levels of granularity is proposed to measure the relevance between utterances and responses directly. We evaluate our proposed approach in both English and Chinese: (i) For English, we evaluate DocChat on WikiQA and QASent two answer sentence selection tasks, and compare it with state-of-the-art methods. Reasonable improvements and good adaptability are observed. (ii) For Chinese, we compare DocChat with Xiaoice, a famous Chinese chitchat engine, and side-by-side evaluation shows that DocChat is a perfect complement for chatbot engines using U-R pairs as main source of responses.

Investigating the Sources of Linguistic Alignment in Conversation*Gabriel Doyle and Michael C. Frank*

15:50–16:10

In conversation, speakers tend to “accommodate” or “align” to their partners, changing the style and substance of their communications to be more similar to their partners’ utterances. We focus here on “linguistic alignment,” changes in word choice based on others’ choices. Although linguistic alignment is observed across many different contexts and its degree correlates with important social factors such as power and likability, its sources are still uncertain. We build on a recent probabilistic model of alignment, using it to separate out alignment attributable to words versus word categories. We model alignment in two contexts: telephone conversations and microblog replies. Our results show evidence of alignment, but it is primarily lexical rather than categorical. Furthermore, we find that discourse acts modulate alignment substantially. This evidence supports the view that alignment is shaped by strategic communicative processes related to the ongoing discourse.

Entropy Converges Between Dialogue Participants: Explanations from an Information-Theoretic Perspective*Yang Xu and David Reitter*

16:10–16:30

The applicability of entropy rate constancy to dialogue is examined on two spoken dialogue corpora. The principle is found to hold; however, new entropy change patterns within the topic episodes of dialogue are described, which are different from written text. Speaker’s dynamic roles as topic initiators and topic responders are associated with decreasing and increasing entropy, respectively, which results in local convergence between these speakers in each topic episode. This implies that the sentence entropy in dialogue is conditioned on different contexts determined by the speaker’s roles. Explanations from the perspectives of grounding theory and interactive alignment are discussed, resulting in a novel, unified information-theoretic approach of dialogue.

Finding the Middle Ground - A Model for Planning Satisficing Answers*Sabine Janzen, Wolfgang Maäß, and Tobias Kowatsch*

16:30–16:50

To establish sophisticated dialogue systems, text planning needs to cope with congruent as well as incongruent interlocutor interests as given in everyday dialogues. Little attention has been given to this topic in text planning in contrast to dialogues that are fully aligned with anticipated user interests. When considering dialogues with congruent and incongruent interlocutor interests, dialogue partners are facing the constant challenge of finding a balance between cooperation and competition. We introduce the concept of fairness that operationalizes an equal and adequate, i.e. equitable satisfaction of all interlocutors’ interests. Focusing on Question-Answering (QA) settings, we describe an answer planning approach that support fair dialogues under congruent and incongruent interlocutor interests. Due to the fact that fairness is subjective per se, we present promising results from an empirical study ($N=107$) in which human subjects interacted with a QA system implementing the proposed approach.

A Sentence Interaction Network for Modeling Dependence between Sentences*Biao Liu and Minlie Huang*

16:50–17:10

Modeling interactions between two sentences is crucial for a number of natural language processing tasks including Answer Selection, Dialogue Act Analysis, etc. While deep learning methods like Recurrent Neural Network or Convolutional Neural Network have been proved to be powerful for sentence modeling, prior studies paid less attention on interactions between sentences. In this work, we propose a Sentence Interaction Network (SIN) for modeling the complex interactions between two sentences. By introducing “interaction states” for word and phrase pairs, SIN is powerful and flexible in capturing sentence interactions for different tasks. We obtain significant improvements on Answer Selection and Dialogue Act Analysis without any feature engineering.

3E: Generation

Chair:

Towards more variation in text generation: Developing and evaluating variation models for choice of referential form

Thiago Castro Ferreira, Emiel Krahmer, and Sander Wubben

15:30–15:50

In this study, we introduce a non-deterministic method for referring expression generation. We describe two models that account for individual variation in the choice of referential form in automatically generated text: a Naïve Bayes model and a Recurrent Neural Network. Both are evaluated using the VaREG corpus. Then we select the best performing model to generate referential forms in texts from the GREC-2.0 corpus and conduct an evaluation experiment in which humans judge the coherence and comprehensibility of the generated texts, comparing them both with the original references and those produced by a random baseline model.

How Much is 131 Million Dollars? Putting Numbers in Perspective with Compositional Descriptions

Arun Chaganty and Percy Liang

15:50–16:10

How much is 131 million US dollars? To help readers put such numbers in context, we propose a new task of automatically generating short descriptions known as perspectives, e.g. “\$131 million is about the cost to employ everyone in Texas over a lunch period”. First, we collect a new dataset of numeric mentions from news articles paired with ratings of generated perspectives using crowdsourcing. We then propose a system to generate these descriptions consisting of two steps: formula construction and description generation. In construction, we compose formulae from numeric facts in a knowledge base and rank the resulting formulas based on familiarity, numeric proximity and semantic compatibility. In generation, we convert a formula into natural language using a sequence-to-sequence neural network model. Our system has a 15.2% F1 improvement over a non-compositional baseline at formula construction and a 12.5 BLEU point improvement over a baseline at description generation.

Generating Factoid Questions With Recurrent Neural Networks: The 30M Factoid Question-Answer Corpus

Iulian Vlad Serban, Alberto García-Durán, Caglar Gulcehre, Sungjin Ahn, Sarath Chandar, Aaron Courville, and Yoshua Bengio

16:10–16:30

Over the past decade, large-scale supervised learning corpora have enabled machine learning researchers to make substantial advances. However, to this date, there are no large-scale question-answer corpora available. In this paper we present the 30M Factoid Question-Answer Corpus, an enormous question-answer pair corpus produced by applying a novel neural network architecture on the knowledge base Freebase to transduce facts into natural language questions. The produced question-answer pairs are evaluated both by human evaluators and using automatic evaluation metrics, including well-established machine translation and sentence similarity metrics. Across all evaluation criteria the question-generation model outperforms the competing template-based baseline. Furthermore, when presented to human evaluators, the generated questions appear to be comparable in quality to real human-generated questions.

Latent Predictor Networks for Code Generation

Wang Ling, Phil Blunsom, Edward Grefenstette, Karl Moritz Hermann, Tomáš Kočiský, Fumin Wang, and Andrew Senior

16:30–16:50

Many language generation tasks require the production of text conditioned on both structured and unstructured inputs. We present a novel neural network architecture which generates an output sequence conditioned on an arbitrary number of input functions. Crucially, our approach allows both the choice of conditioning context and the granularity of generation, for example characters or tokens, to be marginalised, thus permitting scalable and effective training. Using this framework, we address the problem of generating programming code from a mixed natural language and structured specification. We create two new data sets for this paradigm derived from the collectible trading card games Magic the Gathering and Hearthstone. On these, and a third preexisting corpus, we demonstrate that marginalising multiple predictors allows our model to outperform strong benchmarks.

Easy Things First: Installments Improve Referring Expression Generation for Objects in Photographs

Sina Zarrieß and David Schlangen

16:50–17:10

Research on generating referring expressions has so far mostly focussed on “one-shot reference”, where the aim is to generate a single, discriminating expression. In interactive settings, however, it is not uncommon for reference to be established in “installments”, where referring information is offered piecewise until success has been confirmed. We show that this strategy can also be advantageous in technical systems that only have uncertain access to object attributes and categories. We train a recently introduced model of grounded word meaning on a data set of REs for objects in images and learn to predict semantically appropriate expressions. In a human evaluation, we observe that users are sensitive to inadequate object names – which unfortunately are not unlikely to be generated from low-level visual input. We propose a solution inspired from human task-oriented interaction and implement strategies for avoiding and repairing semantically inaccurate words. We enhance a word-based REG with context-aware, referential installments and find that they substantially improve the referential success of the system.

3F: Entities and coreference*Chair:***Collective Entity Resolution with Multi-Focal Attention***Amir Globerson, Nevena Lazic, Soumen Chakrabarti, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira* 15:30–15:50

Entity resolution is the task of linking each mention of an entity in text to the corresponding record in a knowledge base (KB). Coherence models for entity resolution encourage all referring expressions in a document to resolve to entities that are related in the KB. We explore attention-like mechanisms for coherence, where the evidence for each candidate is based on a small set of strong relations, rather than relations to all other entities in the document. The rationale is that document-wide support may simply not exist for non-salient entities, or entities not densely connected in the KB. Our proposed system outperforms state-of-the-art systems on the CoNLL 2003, TAC KBP 2010, 2011 and 2012 tasks.

Online Adaptor Grammars with Hybrid inference*Nevena Lazic, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira* 15:50–16:10

We present Plato, a probabilistic model for entity resolution that includes a novel approach for handling noisy or uninformative features, and supplements labeled training data derived from Wikipedia with a very large unlabeled text corpus. Training and inference in the proposed model can easily be distributed across many servers, allowing it to scale to over 10^7 entities. We evaluate Plato on three standard datasets for entity resolution. We show that it is competitive on both the CoNLL 2003 and TAC KBP 2012 datasets.

Online Adaptor Grammars with Hybrid inference*Dat Ba Nguyen, Martin Theobald, and Gerhard Weikum* 16:10–16:30

Methods for Named Entity Recognition and Disambiguation (NERD) perform NER and NED in two separate stages. Therefore, NED may be penalized with respect to precision by NER false positives, and suffers in recall from NER false negatives. Conversely, NED does not fully exploit information computed by NER such as types of mentions. This paper presents J-NERD, a new approach to perform NER and NED jointly, by means of a probabilistic graphical model that captures mention spans, mention types, and the mapping of mentions to entities in a knowledge base. We present experiments with different kinds of texts from the CoNLL'03, ACE'05, and ClueWeb'09-FACC1 corpora. J-NERD consistently outperforms state-of-the-art competitors in end-to-end NERD precision, recall, and F1.

Which Coreference Evaluation Metric Do You Trust? A Proposal for a Link-based Entity Aware Metric*Nafise Sadat Moosavi and Michael Strube* 16:30–16:50

Interpretability and discriminative power are the two most basic requirements for an evaluation metric. In this paper, we report the mention identification effect in the B3, CEAf, and BLANC coreference evaluation metrics that makes it impossible to interpret their results properly. The only metric which is insensitive to this flaw is MUC, which, however, is known to be the least discriminative metric. It is a known fact that none of the current metrics are reliable. The common practice for ranking coreference resolvers is to use the average of three different metrics. However, one cannot expect to obtain a reliable score by averaging three unreliable metrics. We propose LEA, a Link-based Entity-Aware evaluation metric that is designed to overcome the shortcomings of the current evaluation metrics. LEA is available as branch LEA-scorer in the reference implementation of the official CoNLL scorer.

Improving Coreference Resolution by Learning Entity-Level Distributed Representations*Kevin Clark and Christopher D. Manning* 16:50–17:10

A long-standing challenge in coreference resolution has been the incorporation of entity-level information – features defined over clusters of mentions instead of mention pairs. We present a neural network based coreference system that produces high-dimensional vector representations for pairs of coreference clusters. Using these representations, our system learns when combining clusters is desirable. We train the system with a learning-to-search algorithm that teaches it which local decisions (cluster merges) will lead to a high-scoring final coreference partition. The system substantially outperforms the current state-of-the-art on the English and Chinese portions of the CoNLL 2012 Shared Task dataset despite using few hand-engineered features.

3G: Topics

Chair:

Effects of Creativity and Cluster Tightness on Short Text Clustering Performance

Catherine Finegan-Dollak, Reed Coke, Rui Zhang, Xiangyi Ye, and Dragomir Radev 15:30–15:50

Properties of corpora, such as the diversity of vocabulary and how tightly related texts cluster together, impact the best way to cluster short texts. We examine several such properties in a variety of corpora and track their effects on various combinations of similarity metrics and clustering algorithms. We show that semantic similarity metrics outperform traditional n -gram and dependency similarity metrics for k-means clustering of a linguistically creative dataset, but do not help with less creative texts. Yet the choice of similarity metric interacts with the choice of clustering method. We find that graph-based clustering methods perform well on tightly clustered data but poorly on loosely clustered data. Semantic similarity metrics generate loosely clustered output even when applied to a tightly clustered dataset. Thus, the best performing clustering systems could not use semantic metrics.

Generative Topic Embedding: a Continuous Representation of Documents

Shaohua Li, Tat-Seng Chua, Jun Zhu, and Chunyan Miao

15:50–16:10

Word embedding maps words into a low-dimensional continuous embedding space by exploiting the local word collocation patterns in a small context window. On the other hand, topic modeling maps documents onto a low-dimensional topic space, by utilizing the global word collocation patterns in the same document. These two types of patterns are complementary. In this paper, we propose a generative topic embedding model to combine the two types of patterns. In our model, topics are represented by embedding vectors, and are shared across documents. The probability of each word is influenced by both its local context and its topic. A variational inference method yields the topic embeddings as well as the topic mixing proportions for each document. Jointly they represent the document in a low-dimensional continuous space. In two document classification tasks, our method performs better than eight existing methods, with fewer features. In addition, we illustrate with an example that our method can generate coherent topics even based on only one document.

Detecting Cross-cultural Differences Using a Multilingual Topic Model

E. Dario Gutierrez, Ekaterina Shutova, Patricia Lichtenstein, Gerard de Melo, and Luca Gilardi
16:10–16:30

Understanding cross-cultural differences has important implications for world affairs and many aspects of the life of society. Yet, the majority of text-mining methods to date focus on the analysis of monolingual texts. In contrast, we present a statistical model that simultaneously learns a set of common topics from multilingual, non-parallel data and automatically discovers the differences in perspectives on these topics across linguistic communities. We perform a behavioural evaluation of a subset of the differences identified by our model in English and Spanish to investigate their psychological validity.

Detecting Common Discussion Topics Across Culture From News Reader Comments

Bei Shi, Wai Lam, Lidong Bing, and Yingqin Xu 16:30–16:50

News reader comments found in many on-line news websites are typically massive in amount. We investigate the task of Cultural-common Topic Detection (CTD), which is aimed at discovering common discussion topics from news reader comments written in different languages. We propose a new probabilistic graphical model called MCTA which can cope with the language gap and capture the common semantics in different languages. We also develop a partially collapsed Gibbs sampler which effectively incorporates the term translation relationship into the detection of cultural-common topics for model parameter learning. Experimental results show improvements over the state-of-the-art model.

A Discriminative Topic Model using Document Network Structure

Weiwei Yang, Jordan Boyd-Graber, and Philip Resnik 16:50–17:10

Document collections often have links between documents—citations, hyperlinks, or revisions—and which links are added is often based on topical similarity. To model these intuitions, we introduce a new topic model for documents situated within a network structure, integrating latent blocks of documents with a max-margin learning criterion for link prediction using topic- and word-level features. Experiments on a scientific paper dataset and collection of webpages show that, by more robustly exploiting the rich link structure within a document network, our model improves link prediction, topic quality, and block distributions.

4

Main Conference: Tuesday, June 2

Overview

9:00–10:10	Invited talk II: Mark Steedman	<i>Plaza Ballroom A, B, & C</i>				
10:10–10:40	Coffee break					
Session 4						
10:40–12:00	Relations and knowledge bases	Semantic parsing II	Word meaning II	Tasks and datasets	Parsing IV	Document analysis
12:00–1:40 Lunch break						
Session 5						
1:40–3:00	Deep learning (short papers)	Semantics and generation (short papers)	Machine translation II (short papers)	Text classification (short papers)	Potpourri I (short papers)	
3:00–3:30 Coffee break						
Session 6						
3:30–5:10	Machine learning	Word vectors II	Machine translation III	Discourse	Language and vision	Summarization/hearer language
5:30–7:00	Poster and dinner session II					<i>Plaza Ballroom A, B, & C</i>
6:00–7:30	ACL System Demonstrations Session B					
8:00–11:00	Social event					<i>Grand Ballroom</i>

Session 4 Overview – Tuesday, June 2, 2015

	Track A <i>Relations and knowledge bases</i> Plaza Ballroom A	Track B <i>Semantic parsing II</i> Plaza Ballroom D	Track C <i>Word meaning II</i> Plaza Ballroom F	Track D <i>Tasks and datasets</i>	Track E <i>Parsing IV</i>	Track F <i>Document analysis</i>
10:40	Identifying Causal Relations Using Parallel Wikipedia Articles <i>Hidey and McKeown</i>	Simpler Context-Dependent Logical Forms via Model Projections <i>Long, Pasupat, and Liang</i>		Beyond Plain Spatial Knowledge: Determining Where Entities Are and Are Not Located, and For How Long <i>Vempala and Blanco</i>	Word Embeddings as Metric Recovery in Semantic Spaces <i>Kiperwasser and Goldberg</i>	Inferring Perceived Demographics from User Emotional Tone and User-Environment Emotional Contrast <i>Volkova and Bachrach</i>
11:00	Compositional Learning of Embeddings for Relation Paths in Knowledge Base and Text <i>Toutanova, Lin, Yih, Poon, and Quirk</i>		Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations <i>Frermann and Lapata</i>	LexSemTm: A Semantic Dataset Based on All-words Unsupervised Sense Distribution Learning <i>Bennett, Baldwin, Lau, McCarthy, and Bond</i>	Optimizing Spectral Learning for Parsing <i>Narayan and Cohen</i>	Prototype Synthesis for Model Laws <i>Burgess, Giraudy, and Adar</i>
11:20	Commonsense Knowledge Base Completion <i>Li, Taheri, Tu, and Gimpel</i>	A Fast Unified Model for Parsing and Sentence Understanding <i>Bowman, Gauthier, Rastogi, Gupta, Manning, and Potts</i>	Investigating Language Universal and Specific Properties in Word Embeddings <i>Qian, Qiu, and Huang</i>	The LAMBADA dataset: Word prediction requiring a broad discourse context <i>Paperno, Kruszewski, Lazaridou, Pham, Bernardi, Pezzelle, Baroni, Boleda, and Fernandez</i>	Stack-propagation: Improved Representation Learning for Syntax <i>Zhang and Weiss</i>	Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM <i>Habernal and Gurevych</i>
11:40	Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations <i>Marcheggiani and Titov</i>	Online Adaptor Grammars with Hybrid inference <i>Richardson and Kuhn</i>	Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change <i>Hamilton, Leskovec, and Jurafsky</i>	WikiReading: A Novel Large-scale Language Understanding Task over Wikipedia <i>Hewlett, Lacoste, Jones, Polosukhin, Fandrianto, Han, Kelcey, and Berthelot</i>		Discovery of Treatments from Text Corpora <i>Fong and Grimmer</i>

Parallel Session 4

4A: Relations and knowledge bases

Plaza Ballroom A & B

Chair:

Identifying Causal Relations Using Parallel Wikipedia Articles

Christopher Hidey and Kathy McKeown

10:40–11:00

The automatic detection of causal relationships in text is important for natural language understanding. This task has proven to be difficult, however, due to the need for world knowledge and inference. We focus on a sub-task of this problem where an open class set of linguistic markers can provide clues towards understanding causality. Unlike the explicit markers, a closed class, these markers vary significantly in their linguistic forms. We leverage parallel Wikipedia corpora to identify new markers that are variations on known causal phrases, creating a training set via distant supervision. We also train a causal classifier using features from the open class markers and semantic features providing contextual information. The results show that our features provide an 11.05 point absolute increase over the baseline on the task of identifying causality in text.

Compositional Learning of Embeddings for Relation Paths in Knowledge Base and Text

Kristina Toutanova, Victoria Lin, Wen-tau Yih, Hoifung Poon, and Chris Quirk

11:00–11:20

Modeling relation paths has been shown to offer significant gains in embedding models for knowledge base (KB) completion. However, enumerating paths between two entities is very expensive, and existing approaches typically resort to an approximation using a sampled subset. This problem is particularly acute when text is jointly modeled with KB relations and used to provide direct evidence for facts mentioned in it. In this paper, we propose the first exact dynamic programming algorithm which enables efficient incorporation of all multi-step relation paths of bounded length, while also modeling both relation types and intermediate nodes in the compositional path representations. We conduct a theoretical analysis on the efficiency gain from such an approach. Experiments on biomedical domain show that it addresses representational limitations in prior approaches and improves accuracy in KB completion.

Commonsense Knowledge Base Completion

Xiang Li, Aynaz Taheri, Lifu Tu, and Kevin Gimpel

11:20–11:40

We enrich a curated resource of commonsense knowledge by formulating the problem as one of knowledge base completion (KBC). Most work in KBC focuses on knowledge bases like Freebase that relate entities drawn from a fixed set. However, the tuples in ConceptNet (Speer and Havasi, 2012) define relations between an unbounded set of phrases. We develop neural network models for scoring tuples on arbitrary phrases and evaluate them by their ability to distinguish true held-out tuples from false ones. We find strong performance from a bilinear model using a simple additive architecture to model phrases. We manually evaluate our trained model’s ability to assign quality scores to novel tuples, finding that it can propose tuples at the same quality level as medium-confidence tuples from ConceptNet.

Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations

Diego Marcheggiani and Ivan Titov

11:40–12:00

We present a method for unsupervised open-domain relation discovery. In contrast to previous (mostly generative and agglomerative clustering) approaches, our model relies on rich contextual features and makes minimal independence assumptions. The model is composed of two parts: a feature-rich relation extractor, which predicts a semantic relation between two entities, and a factorization model, which reconstructs arguments (i.e., the entities) relying on the predicted relation. The two components are estimated jointly so as to minimize errors in recovering arguments. We study factorization models inspired by previous work in relation factorization and selectional preference modeling. Our models substantially outperform the generative and agglomerative-clustering counterparts and achieve state-of-the-art performance.

4B: Semantic parsing II

Plaza Ballroom D & E

Chair:

Simpler Context-Dependent Logical Forms via Model Projections

Reginald Long, Panupong Pasupat, and Percy Liang

10:40–11:00

We consider the task of learning a context-dependent mapping from utterances to denotations. With only denotations at training time, we must search over a combinatorially large space of logical forms, which is even larger with context-dependent utterances. To cope with this challenge, we perform successive projections of the full model onto simpler models that operate over equivalence classes of logical forms. Though less expressive, we find that these simpler models are much faster and can be surprisingly effective. Moreover, they can be used to bootstrap the full model. Finally, we collected three new context-dependent semantic parsing datasets, and develop a new left-to-right parser.

A Fast Unified Model for Parsing and Sentence Understanding

Samuel R. Bowman, Jon Gauthier, Abhinav Rastogi, Raghav Gupta, Christopher D. Manning, and Christopher Potts

11:20–11:40

Tree-structured neural networks exploit valuable syntactic parse information as they interpret the meanings of sentences. However, they suffer from two key technical problems that make them slow and unwieldy for large-scale NLP tasks: they usually operate on parsed sentences and they do not directly support batched computation. We address these issues by introducing the Stack-augmented Parser-Interpreter Neural Network (SPINN), which combines parsing and interpretation within a single tree-sequence hybrid model by integrating tree-structured sentence interpretation into the linear sequential structure of a shift-reduce parser. Our model supports batched computation for a speedup of up to 25 times over other tree-structured models, and its integrated parser can operate on unparsed data with little loss in accuracy. We evaluate it on the Stanford NLI entailment task and show that it significantly outperforms other sentence-encoding models.

Online Adaptor Grammars with Hybrid inference

Kyle Richardson and Jonas Kuhn

11:40–12:00

We introduce a new approach to training a semantic parser that uses textual entailment judgements as supervision. These judgements are based on high-level inferences about whether the meaning of one sentence follows from another. When applied to an existing semantic parsing task, they prove to be a useful tool for revealing semantic distinctions and background knowledge not captured in the target representations. This information is used to improve the quality of the semantic representations being learned and to acquire generic knowledge for reasoning. Experiments are done on the benchmark Sportscaster corpus (Chen and Mooney, 2008), and a novel RTE-inspired inference dataset is introduced. On this new dataset our method strongly outperforms several strong baselines. Separately, we obtain state-of-the-art results on the original Sportscaster semantic parsing task.

4C: Word meaning II

Plaza Ballroom F

Chair:

Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations
Lea Freymann and Mirella Lapata

11:00–11:20

Word meanings change over time and an automated procedure for extracting this information from text would be useful for historical exploratory studies, information retrieval or question answering. We present a dynamic Bayesian model of diachronic meaning change, which infers temporal word representations as a set of senses and their prevalence. Unlike previous work, we explicitly model language change as a smooth, gradual process. We experimentally show that this modeling decision is beneficial: our model performs competitively on meaning change detection tasks whilst inducing discernible word senses and their development over time. Application of our model to the SemEval-2015 temporal classification benchmark datasets further reveals that it performs on par with highly optimized task-specific systems.

Investigating Language Universal and Specific Properties in Word Embeddings*Peng Qian, Xipeng Qiu, and Xuanjing Huang*

11:20–11:40

Recently, many NLP tasks have benefited from distributed word representation. However, it remains unknown whether embedding models are really immune to the typological diversity of languages, despite the language-independent architecture. Here we investigate three models on a large set of language samples by mapping dense embedding to sparse linguistic property space. Experiment results reveal the language universal and specific encoded in various word representation. Additionally, strong evidence supports the utility of word form, especially for inflectional languages.

Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change*William L. Hamilton, Jure Leskovec, and Dan Jurafsky*

11:40–12:00

Understanding how words change their meanings over time is key to models of language and cultural evolution, but historical data on meaning is scarce, making theories hard to develop and test. Word embeddings show promise as a diachronic tool, but have not been carefully evaluated. We develop a robust methodology for quantifying semantic change by evaluating word embeddings (PPMI, SVD, word2vec) against known historical changes. We then use this methodology to reveal statistical laws of semantic evolution. Using six historical corpora spanning four languages and two centuries, we propose two quantitative laws of semantic change: (i) the law of conformity—the rate of semantic change scales with an inverse power-law of word frequency; (ii) the law of innovation—Independent of frequency, words that are more polysemous have higher rates of semantic change.

4D: Tasks and datasets

Chair:

Beyond Plain Spatial Knowledge: Determining Where Entities Are and Are Not Located, and For How Long

Alakananda Vempala and Eduardo Blanco

10:40–11:00

This paper complements semantic role representations with spatial knowledge beyond indicating plain locations. Namely, we extract where entities are (and are not) located, and for how long (seconds, hours, days, etc.). Crowdsourced annotations show that this additional knowledge is intuitive to humans and can be annotated by non-experts. Experimental results show that the task can be automated.

LexSemTm: A Semantic Dataset Based on All-words Unsupervised Sense Distribution Learning

Andrew Bennett, Timothy Baldwin, Jey Han Lau, Diana McCarthy, and Francis Bond 11:00–11:20

There has recently been a lot of interest in unsupervised methods for learning sense distributions, particularly in applications where sense distinctions are needed. This paper analyses a state-of-the-art method for sense distribution learning, and optimises it for application to the entire vocabulary of a given language. The optimised method is then used to produce LexSemTM: a sense frequency and semantic dataset of unprecedented size, spanning approximately 88% of polysemous, English simplex lemmas, which is released as a public resource to the community. Finally, the quality of this data is investigated, and the LexSemTM sense distributions are shown to be superior to those based on the WordNet first sense for lemmas missing from SEMCOR, and at least on par with SEMCOR-based distributions otherwise.

The LAMBADA dataset: Word prediction requiring a broad discourse context

Denis Paperno, Germán Kruszewski, Angeliki Lazaridou, Ngoc Quan Pham, Raffaella Bernardi, Sandro Pezzelle, Marco Baroni, Gemma Boleda, and Raquel Fernandez 11:20–11:40

We introduce LAMBADA, a dataset to evaluate the capabilities of computational models for text understanding by means of a word prediction task. LAMBADA is a collection of narrative passages sharing the characteristic that human subjects are able to guess their last word if they are exposed to the whole passage, but not if they only see the last sentence preceding the target word. To succeed on LAMBADA, computational models cannot simply rely on local context, but must be able to keep track of information in the broader discourse. We show that LAMBADA exemplifies a wide range of linguistic phenomena, and that none of several state-of-the-art language models reaches accuracy above 1% on this novel benchmark. We thus propose LAMBADA as a challenging test set, meant to encourage the development of new models capable of genuine understanding of broad context in natural language text.

WikiReading: A Novel Large-scale Language Understanding Task over Wikipedia

Daniel Hewlett, Alexandre Lacoste, Llion Jones, Illia Polosukhin, Andrew Fardianto, Jay Han, Matthew Kelcey, and David Berthelot 11:40–12:00

We present WikiReading, a large-scale natural language understanding task and publicly-available dataset with 18 million instances. The task is to predict textual values from the structured knowledge base Wikidata by reading the text of the corresponding Wikipedia articles. The task contains a rich variety of challenging classification and extraction sub-tasks, making it well-suited to end-to-end models such as deep neural networks (DNNs). We compare various state-of-the-art DNN-based architectures for document classification, information extraction, and question answering. We find that models supporting a rich answer space, such as word or character sequences, perform best. Our best-performing model, a word-level sequence to sequence model with a mechanism to copy out-of-vocabulary words, obtains an accuracy of 71.8%.

4E: Parsing IV*Chair:***Word Embeddings as Metric Recovery in Semantic Spaces***Elyahu Kiperwasser and Yoav Goldberg*

10:40–11:00

We present a simple and effective scheme for dependency parsing which is based on bidirectional-LSTMs (BiLSTMs). Each sentence token is associated with a BiLSTM vector representing the token in its sentential context, and feature vectors are constructed by concatenating a few BiLSTM vectors. The BiLSTM is trained jointly with the parser objective, resulting in very effective feature extractors for parsing. We demonstrate the effectiveness of the approach by applying it to a greedy transition based parser as well as to a globally optimized graph-based parser. The resulting parsers have very simple architectures, and match or surpass the state-of-the-art accuracies on English and Chinese.

Optimizing Spectral Learning for Parsing*Shashi Narayan and Shay B. Cohen*

11:00–11:20

We describe a search algorithm for optimizing the number of latent states when estimating latent-variable PCFGs with spectral methods. Our results show that contrary to the common belief that the number of latent states for each nonterminal in an L-PCFG can be decided in isolation with spectral methods, parsing results significantly improve if the number of latent states for each nonterminal is globally optimized, while taking into account interactions between the different nonterminals. In addition, we contribute an empirical analysis of spectral algorithms on eight morphologically rich languages: Basque, French, German, Hebrew, Hungarian, Korean, Polish and Swedish. Our results show that our estimation consistently performs better or close to coarse-to-fine expectation-maximization techniques for these languages.

Stack-propagation: Improved Representation Learning for Syntax*Yuan Zhang and David Weiss*

11:20–11:40

Traditional syntax models typically leverage part-of-speech (POS) information by constructing features from hand-tuned templates. We demonstrate that a better approach is to utilize POS tags as a *regularizer* of learned representations. We propose a simple method for learning a stacked pipeline of models which we call “stack-propagation.” We apply this to dependency parsing and tagging, where we use the hidden layer of the tagger network as a representation of the input tokens for the parser. At test time, our parser does not require predicted POS tags. On 19 languages from the Universal Dependencies, our method is 1.3% (absolute) more accurate than a state-of-the-art graph-based approach and 2.7% more accurate than the most comparable greedy model.

4F: Document analysis

Chair:

Inferring Perceived Demographics from User Emotional Tone and User-Environment Emotional Contrast

Svitlana Volkova and Yoram Bachrach

10:40–11:00

We examine communications in a social network to study user emotional contrast - the propensity of users to express different emotions than those expressed by their neighbors. Our analysis is based on a large Twitter dataset, consisting of the tweets of 123,513 users from the USA and Canada. Focusing on Ekman's basic emotions, we analyze differences between the emotional tone expressed by these users and their neighbors and correlate these differences with perceived user demographics. We demonstrate that many perceived user demographic traits correlate with the emotional contrast between users and their neighbors. Unlike other approaches on inferring user demographics in social media that rely solely on user communications, we explore the network structure and show that it is possible to accurately predict a range of perceived demographic traits based solely on the emotions emanating from users and their neighbors.

Prototype Synthesis for Model Laws

Matthew Burgess, Eugenia Giraudy, and Eytan Adar

11:00–11:20

State legislatures often rely on existing text when drafting new bills. Resource and expertise constraints, which often drive this copying behavior, can be taken advantage of by lobbyists and special interest groups. These groups provide model bills, which encode policy agendas, with the intent that the models become actual law. Unfortunately, model legislation is often opaque to the public—both in source and content. In this paper we present LOBBYBACK, a system that reverse engineers model legislation from observed text. LOBBYBACK identifies clusters of bills which have text reuse and generates “prototypes” that represent a canonical version of the text shared between the documents. We demonstrate that LOBBYBACK accurately reconstructs model legislation and apply it to a dataset of over 550k bills.

Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM

Ivan Habernal and Iryna Gurevych

11:20–11:40

We propose a new task in the field of computational argumentation in which we investigate qualitative properties of Web arguments, namely their convincingness. We cast the problem as relation classification, where a pair of arguments having the same stance to the same prompt is judged. We annotate a large datasets of 16k pairs of arguments over 32 topics and investigate whether the relation “A is more convincing than B” exhibits properties of total ordering; these findings are used as global constraints for cleaning the crowdsourced data. We propose two tasks: (1) predicting which argument from an argument pair is more convincing and (2) ranking all arguments to the topic based on their convincingness. We experiment with feature-rich SVM and bidirectional LSTM and obtain 0.76–0.78 accuracy and 0.35–0.40 Spearman’s correlation in a cross-topic evaluation. We release the newly created corpus UKPConvArg¹ and the experimental software under open licenses.

Discovery of Treatments from Text Corpora

Christian Fong and Justin Grimmer

11:40–12:00

An extensive literature in computational social science examines how features of messages, advertisements, and other corpora affect individuals’ decisions, but these analyses must specify the relevant features of the text before the experiment. Automated text analysis methods are able to discover features of text, but these methods cannot be used to obtain the estimates of causal effects—the quantity of interest for applied researchers. We introduce a new experimental design and statistical model to simultaneously discover treatments in a corpora and estimate causal effects for these discovered treatments. We prove the conditions to identify the treatment effects of texts and introduce the supervised Indian Buffet process to discover those treatments. Our method enables us to discover treatments in a training set using a collection of texts and individuals’ responses to those texts, and then estimate the effects of these interventions in a test set of new texts and survey respondents. We apply the model to an experiment about candidate biographies, recovering intuitive features of voters’ decisions and revealing a penalty for lawyers and a bonus for military service.

Session 5 Overview – Tuesday, June 2, 2015

Track A <i>Deep learning (short papers)</i>	Track B <i>Semantics and generation (short papers)</i>	Track C <i>Machine translation II (short papers)</i>	Track D <i>Text classification (short papers)</i>	Track E <i>Potpourri I (short papers)</i>
Plaza Ballroom A &	Plaza Ballroom D &	Plaza Ballroom F		
The red one!: On learning to refer to things based on discriminative properties <i>Lazaridou, Pham, and Baroni</i>	An Entity-Focused Approach to Generating Company Descriptions <i>Saldanha, Biran, McKeown, and Gliozzo</i>	Phrase-Level Combination of SMT and TM Using Constrained Word Lattice <i>Li, Way, and Liu</i>	Very quaffable and great fun: Applying NLP to wine reviews <i>Hendrickx, Lefever, Croijmans, Majid, and Bosch</i>	Text Simplification as Tree Labeling <i>Bingel and Søgaard</i>
Don't Count, Predict! An Automatic Approach to Learning Sentiment Lexicons for Short Text <i>Vo and Zhang</i>	Annotating Relation Inference in Context via Question Answering <i>Levy and Dagan</i>	A Neural Network based Approach to Automatic Post-Editing <i>Pal, Naskar, Vela, and Genabith</i>	Exploring Stylistic Variation with Age and Income on Twitter <i>Flekova, Preoțiuc-Pietro, and Ungar</i>	Bootstrapped Text-level Named Entity Recognition for Literature <i>Brooke, Hammond, and Baldwin</i>
Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model <i>Wang, Yu, Lai, and Zhang</i>	Automatic Semantic Classification of German Preposition Types: Comparing Hard and Soft Clustering Approaches across Features <i>Köper and Schulte im Walde</i>	An Unsupervised Method for Automatic Translation Memory Cleaning <i>Jalili Sabet, Negri, Turchi, and Barbu</i>	Finding Optimists and Pessimists on Twitter <i>Ruan, Wilson, and Mihalcea</i>	The Enemy in Your Own Camp: How Well Can We Detect Statistically-Generated Fake Reviews — An Adversarial Study <i>Hovy</i>
Deep multi-task learning with low level tasks supervised at lower layers <i>Søgaard and Goldberg</i>	Natural Language Generation enhances human decision-making with uncertain information <i>Gkatzia, Lemon, and Rieser</i>	Exponentially Decaying Bag-of-Words Input Features for Feed-Forward Neural Network in Statistical Machine Translation <i>Peter, Wang, and Ney</i>	Transductive Adaptation of Black Box Predictions <i>Clinchant, Chidlovskii, and Csurka</i>	
Domain Specific Named Entity Recognition Referring to the Real World by Deep Neural Networks <i>Tomori, Ninomiya, and Mori</i>	Tweet2Vec: Character-Based Distributed Representations for Social Media <i>Dhangra, Zhou, Fitzpatrick, Muehl, and Cohen</i>	Syntactically Guided Neural Machine Translation <i>Stahlberg, Hasler, Waite, and Byrne</i>	Which Tumblr Post Should I Read Next? <i>Kozareva and Yamada</i>	

01:40

01:56

02:12

02:28

02:44

Parallel Session 5

5A: Deep learning (short papers)

Plaza Ballroom A & B

Chair:

The red one!: On learning to refer to things based on discriminative properties

Angeliki Lazaridou, Nghia The Pham, and Marco Baroni

13:40–13:56

As a first step towards agents learning to communicate about their visual environment, we propose a system that, given visual representations of a referent (cat) and a context (sofa), identifies their discriminative attributes, i.e., properties that distinguish them (has_tail). Moreover, despite the lack of direct supervision at the attribute level, the model learns to assign plausible attributes to objects (sofa-has_cushion). Finally, we present a preliminary experiment confirming the referential success of the predicted discriminative attributes.

Don't Count, Predict! An Automatic Approach to Learning Sentiment Lexicons for Short Text

Duy Tin Vo and Yue Zhang

13:56–14:12

We describe an efficient neural network method to automatically learn sentiment lexicons without relying on any manual resources. The method takes inspiration from the NRC method, which gives the best results in SemEval13 by leveraging emoticons in large tweets, using the PMI between words and tweet sentiments to define the sentiment attributes of words. We show that better lexicons can be learned by using them to predict the tweet sentiment labels. By using a very simple neural network, our method is fast and can take advantage of the same data volume as the NRC method. Experiments show that our lexicons give significantly better accuracies on multiple languages compared to the current best methods.

Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model

Jin Wang, Liang-Chih Yu, K. Robert Lai, and Xuejie Zhang

14:12–14:28

Dimensional sentiment analysis aims to recognize continuous numerical values in multiple dimensions such as the valence-arousal (VA) space. Compared to the categorical approach that focuses on sentiment classification such as binary classification (i.e., positive and negative), the dimensional approach can provide more fine-grained sentiment analysis. This study proposes a regional CNN-LSTM model consisting of two parts: regional CNN and LSTM to predict the VA ratings of texts. Unlike a conventional CNN which considers a whole text as input, the proposed regional CNN uses an individual sentence as a region, dividing an input text into several regions such that the useful affective information in each region can be extracted and weighted according to their contribution to the VA prediction. Such regional information is sequentially integrated across regions using LSTM for VA prediction. By combining the regional CNN and LSTM, both local (regional) information within sentences and long-distance dependency across sentences can be considered in the prediction process. Experimental results show that the proposed method outperforms lexicon-based, regression-based, and NN-based methods proposed in previous studies.

Deep multi-task learning with low level tasks supervised at lower layers

Anders Søgaard and Yoav Goldberg

14:28–14:44

In all previous work on deep multi-task learning we are aware of, all task supervisions are on the same (outermost) layer. We present a multi-task learning architecture with deep bi-directional RNNs, where different tasks supervision can happen at different layers. We present experiments in syntactic chunking and CCG supertagging, coupled with the additional task of POS-tagging. We show that it is consistently better to have POS supervision at the innermost rather than the outermost layer. We argue that this is because “low-level” tasks are better kept at the lower layers, enabling the higher-level tasks make use of the shared representation of the lower-level tasks. Finally, we also show how this architecture can be used for domain adaptation.

Domain Specific Named Entity Recognition Referring to the Real World by Deep Neural Networks

Suzushi Tomori, Takashi Ninomiya, and Shinsuke Mori

14:44–15:00

In this paper, we propose a method for referring to the real world to improve named entity recognition (NER) specialized for a domain. Our method adds a stacked auto-encoder to a text-based deep neural network for NER. We first train the stacked auto-encoder only from the real world information, then the entire deep neural network from sentences annotated with NEs and accompanied by real world information. In our experiments, we took Japanese chess as the example. The dataset consists of pairs of a game state and commentary sentences about it annotated with game-specific NE tags. We conducted NER experiments and showed that referring to the real world improves the NER accuracy.

5B: Semantics and generation (short papers)

Plaza Ballroom D & E

Chair:

An Entity-Focused Approach to Generating Company Descriptions*Gavin Saldanha, Or Biran, Kathleen McKeown, and Alfio Gliozzo*

13:40–13:56

Finding quality descriptions on the web, such as those found in Wikipedia articles, of newer companies can be difficult: search engines show many pages with varying relevance, while multi-document summarization algorithms find it difficult to distinguish between core facts and other information such as news stories. In this paper, we propose an entity-focused, hybrid generation approach to automatically produce descriptions of previously unseen companies, and show that it outperforms a strong summarization baseline.

Annotating Relation Inference in Context via Question Answering*Omer Levy and Ido Dagan*

13:56–14:12

We present a new annotation method for collecting data on relation inference in context. We convert the inference task to one of simple factoid question answering, allowing us to easily scale up to 16,000 high-quality examples. Our method corrects a major bias in previous evaluations, making our dataset much more realistic.

Automatic Semantic Classification of German Preposition Types: Comparing Hard and Soft Clustering Approaches across Features*Maximilian Köper and Sabine Schulze im Walde*

14:12–14:28

This paper addresses an automatic classification of preposition types in German, comparing hard and soft clustering approaches and various window- and syntax-based co-occurrence features. We show that (i) the semantically most salient preposition features (i.e., subcategorised nouns) are the most successful, and that (ii) soft clustering approaches are required for the task but reveal quite different attitudes towards predicting ambiguity.

Natural Language Generation enhances human decision-making with uncertain information*Dimitra Gkatzia, Oliver Lemon, and Verena Rieser*

14:28–14:44

Decision-making is often dependent on uncertain data, e.g. data associated with confidence scores or probabilities. We present a comparison of different information presentations for uncertain data and, for the first time, measure their effects on human decision-making. We show that the use of Natural Language Generation (NLG) improves decision-making under uncertainty, compared to state-of-the-art graphical-based representation methods. In a task-based study with 442 adults, we found that presentations using NLG lead to 24% better decision-making on average than the graphical presentations, and to 44% better decision-making when NLG is combined with graphics. We also show that women achieve significantly better results when presented with NLG output (an 87% increase on average compared to graphical presentations).

Tweet2Vec: Character-Based Distributed Representations for Social Media*Bhuwan Dhingra, Zhong Zhou, Dylan Fitzpatrick, Michael Muehl, and William Cohen*

14:44–

15:00

Text from social media provides a set of challenges that can cause traditional NLP approaches to fail. Informal language, spelling errors, abbreviations, and special characters are all commonplace in these posts, leading to a prohibitively large vocabulary size for word-level approaches. We propose a character composition model, tweet2vec, which finds vector-space representations of whole tweets by learning complex, non-local dependencies in character sequences. The proposed model outperforms a word-level baseline at predicting user-annotated hashtags associated with the posts, doing significantly better when the input contains many out-of-vocabulary words or unusual character sequences. Our tweet2vec encoder is publicly available.

5C: Machine translation II (short papers)

Plaza Ballroom F

*Chair:***Phrase-Level Combination of SMT and TM Using Constrained Word Lattice***Liangyou Li, Andy Way, and Qun Liu*

13:40–13:56

Constrained translation has improved statistical machine translation (SMT) by combining it with translation memory (TM) at sentence-level. In this paper, we propose using a constrained word lattice, which encodes input phrases and TM constraints together, to combine SMT and TM at phrase-level. Experiments on English-Chinese and English-French show that our approach is significantly better than previous combination methods, including sentence-level constrained translation and a recent phrase-level combination.

A Neural Network based Approach to Automatic Post-Editing*Santanu Pal, Sudip Kumar Naskar, Mihaela Vela, and Josef van Genabith*

13:56–14:12

We present a neural network based automatic post-editing (APE) system to improve raw machine translation (MT) output. Our neural model of APE (NNAPE) is based on a bidirectional recurrent neural network (RNN) model and consists of an encoder that encodes an MT output into a fixed-length vector from which a decoder provides a post-edited (PE) translation. APE translations produced by NNAPE showed statistically significant improvements of 3.96, 2.68 and 1.35 BLEU points over the original MT, phrase-based APE and hierarchical APE outputs, respectively. Furthermore, human evaluation also revealed that the NNAPE generated PE translations are much better than the original MT output.

An Unsupervised Method for Automatic Translation Memory Cleaning*Masoud Jalili Sabet, Matteo Negri, Marco Turchi, and Eduard Barbu*

14:12–14:28

We address the problem of automatically cleaning a large-scale Translation Memory (TM) in a fully unsupervised fashion, i.e. without human-labelled data. We approach the task by: i) designing a set of features that capture the similarity between two text segments in different languages, ii) use them to induce reliable training labels for a subset of the translation units (TUs) contained in the TM, and iii) use the automatically labelled data to train an ensemble of binary classifiers. We apply our method to clean a test set composed of 1,000 TUs randomly extracted from the English-Italian version of MyMemory, the world's largest public TM. Our results show competitive performance not only against a strong baseline that exploits machine translation, but also against a state-of-the-art method that relies on human-labelled data.

Exponentially Decaying Bag-of-Words Input Features for Feed-Forward Neural Network in Statistical Machine Translation*Jan-Thorsten Peter, Weiyue Wang, and Hermann Ney*

14:28–14:44

Recently, neural network models have achieved consistent improvements in statistical machine translation. However, most networks only use one-hot encoded input vectors of words as their input. In this work, we investigated the exponentially decaying bag-of-words input features for feed-forward neural network translation models and proposed to train the decay rates along with other weight parameters. This novel bag-of-words model improved our phrase-based state-of-the-art system, which already includes a neural network translation model, by up to 0.5% Bleu and 0.6% Ter on three different translation tasks and even achieved a similar performance to the bidirectional LSTM translation model.

Syntactically Guided Neural Machine Translation*Felix Stahlberg, Eva Hasler, Aurelien Waite, and Bill Byrne*

14:44–15:00

We investigate the use of hierarchical phrase-based SMT lattices in end-to-end neural machine translation (NMT). Weight pushing transforms the Hiero scores for complete translation hypotheses, with the full translation grammar score and full n-gram language model score, into posteriors compatible with NMT predictive probabilities. With a slightly modified NMT beam-search decoder we find gains over both Hiero and NMT decoding alone, with practical advantages in extending NMT to very large input and output vocabularies.

5D: Text classification (short papers)*Chair:***Very quaffable and great fun: Applying NLP to wine reviews***Iris Hendrickx, Els Lefever, Ilja Croijmans, Asifa Majid, and Antal van den Bosch* 13:40–13:56

We automatically predict properties of wines on the basis of smell and flavor descriptions from experts' wine reviews. We show wine experts are capable of describing their smell and flavor experiences in wine reviews in a sufficiently consistent manner, such that we can use their descriptions to predict properties of a wine based solely on language. The experimental results show promising F-scores when using lexical and semantic information to predict the color, grape variety, country of origin, and price of a wine. This demonstrates, contrary to popular opinion, that wine experts' reviews really are informative.

Exploring Stylistic Variation with Age and Income on Twitter*Lucie Flekova, Daniel Preotiuc-Pietro, and Lyle Ungar*

13:56–14:12

Writing style allows NLP tools to adjust to the traits of an author. In this paper, we explore the relation between stylistic and syntactic features and authors' age and income. We confirm our hypothesis that for numerous feature types writing style is predictive of income even beyond age. We analyze the predictive power of writing style features in a regression task on two data sets of around 5,000 Twitter users each. Additionally, we use our validated features to study daily variations in writing style of users from distinct income groups. Temporal stylistic patterns not only provide novel psychological insight into user behavior, but are useful for future research and applications in social media.

Finding Optimists and Pessimists on Twitter*Xianzhi Ruan, Steven Wilson, and Rada Mihalcea*

14:12–14:28

Optimism is linked to various personality factors as well as both psychological and physical health, but how does it relate to the way a person tweets? We analyze the online activity of a set of Twitter users in order to determine how well machine learning algorithms can detect a person's outlook on life by reading their tweets. A sample of tweets from each user is manually annotated in order to establish ground truth labels, and classifiers are trained to distinguish between optimistic and pessimistic users. Our results suggest that the words in people's tweets provide ample evidence to identify them as optimists, pessimists, or somewhere in between. Additionally, several applications of these trained models are explored.

Transductive Adaptation of Black Box Predictions*Stephane Clinchant, Boris Chidlovskii, and Gabriela Csurka*

14:28–14:44

Access to data is critical to any machine learning component aimed at training an accurate predictive model. In reality, data is often a subject of technical and legal constraints. Data may contain sensitive topics and data owners are often reluctant to share them. Instead of access to data, they make available decision making procedures to enable predictions on new data. Under the black box classifier constraint, we build an effective domain adaptation technique which adapts classifier predictions in a transductive setting. We run experiments on text categorization datasets and show that significant gains can be achieved, especially in the unsupervised case where no labels are available in the target domain.

Which Tumblr Post Should I Read Next?*Zornitsa Kozareva and Makoto Yamada*

14:44–15:00

Microblogging sites have emerged as major platforms for bloggers to create and consume posts as well as to follow other bloggers and get informed of their updates. Due to the large number of users, and the huge amount of posts they create, it becomes extremely difficult to identify relevant and interesting blog posts. In this paper, we propose a novel convex collective matrix completion (CCMC) method that effectively utilizes user-item matrix and incorporates additional user activity and topic-based signals to recommend relevant content. The key advantage of CCMC over existing methods is that it can obtain a globally optimal solution and can easily scale to large-scale matrices using Hazan's algorithm. To the best of our knowledge, this is the first work which applies and studies CCMC as a recommendation method in social media. We conduct a large scale study and show significant improvement over existing state-of-the-art approaches.

5E: Potpourri I (short papers)

Chair:

Text Simplification as Tree Labeling

Joachim Bingel and Anders Søgaard

13:40–13:56

We present a new, structured approach to text simplification using conditional random fields over top-down traversals of dependency graphs that jointly predicts possible compressions and paraphrases. Our model reaches readability scores comparable to word-based compression approaches across a range of metrics and human judgements while maintaining more of the important information.

Bootstrapped Text-level Named Entity Recognition for Literature

Julian Brooke, Adam Hammond, and Timothy Baldwin

13:56–14:12

We present a named entity recognition (NER) system for tagging fiction: LitNER. Relative to more traditional approaches, litNER has two important properties: (1) it makes use of no hand-tagged data or gazetteers, instead bootstrapping from term clusters; and (2) it leverages multiple instances of the same name in a text. Our experiments show it to substantially outperform off-the-shelf supervised NER systems.

The Enemy in Your Own Camp: How Well Can We Detect Statistically-Generated Fake Reviews — An Adversarial Study

Dirk Hovy

14:12–14:28

Online reviews are a growing market, but it is struggling with fake reviews. They undermine both the value of reviews to the user, and their trust in the review sites. However, fake positive reviews can boost a business, and so a small industry producing fake reviews has developed. The two sides are facing an arms race that involves more and more natural language processing (NLP). So far, NLP has been used mostly for detection, and works well on human-generated reviews. But what happens if NLP techniques are used to generate fake reviews as well? We investigate the question in an adversarial setup, by assessing the detectability of different fake-review generation strategies. We use generative models to produce reviews based on meta-information, and evaluate their effectiveness against deception-detection models and human judges. We find that meta-information helps detection, but that NLP-generated reviews conditioned on such information are also much harder to detect than conventional ones.

Session 6 Overview – Tuesday, June 2, 2015

Track A <i>Machine learning</i> Plaza Ballroom	Track B <i>Word vectors II</i> Ritz Ballroom	Track C <i>Machine translation III</i> D&E Ballroom	Track D <i>Discourse</i> F	Track E <i>Language and vision</i>	Track F <i>Summarization</i>	Track G <i>Learner language</i>
Learning Structured Predictors from Bandit Feedback for Interactive NLP <i>Sokolov, Kreutzer, Lo, and Riezler</i>	Online Adaptor Grammars with Hybrid inference <i>Arora, Li, Liang, Ma, and Risteski</i>	Minimum Risk Training for Neural Machine Translation <i>Shen, Cheng, He, He, Wu, Sun, and Liu</i>	Implicit Discourse Relation Detection via a Deep Architecture with Gated Relevance Network <i>Chen, Zhang, Liu, Qiu, and Huang</i>	Learning Prototypical Event Structure from Photo Albums <i>Boselut, Chen, Warren, Hajishirzi, and Choi</i>		Phrase Structure Annotation and Parsing for Learner English <i>Nagata and Sakaguchi</i>
Online Adaptor Grammars with Hybrid inference <i>Hartmann, Eckle-Kohler, and Gurevych</i>	Morphological Smoothing and Extrapolation of Word Embeddings <i>Cotterell, Schütze, and Eisner</i>	Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations <i>Cuong, Sima'an, and Titov</i>	Model Architectures for Quotation Detection <i>Scheible, Klinger, and Padó</i>	Cross-Lingual Image Caption Generation <i>Miyazaki and Shimizu</i>		A Trainable Spaced Repetition Model for Language Learning <i>Settles and Meeder</i>
Deep Reinforcement Learning with a Natural Language Action Space <i>He, Chen, He, Gao, Li, Deng, and Ostendorf</i>	Cross-lingual Models of Word Embeddings: An Empirical Comparison <i>Upadhyay, Faruqui, Dyer, and Roth</i>	A Character-level Decoder without Explicit Segmentation for Neural Machine Translation <i>Chung, Cho, and Bengio</i>	Speech Act Modeling of Written Asynchronous Conversations with Task-Specific Embeddings and Conditional Structured Models <i>Joty and Hoque</i>	Learning Concept Taxonomies from Multi-modal Data <i>Zhang, Hu, Deng, Sachan, Yan, and Xing</i>		Reassessing the Goals of Grammatical Error Correction: Fluency Instead of Grammaticality <i>Sakaguchi, Napoles, Post, and Tetreault</i>
Incorporating Copying Mechanism in Sequence-to-Sequence Learning <i>Gu, Lu, Li, and Li</i>	Take and Took, Gaggle and Goose, Book and Read: Evaluating the Utility of Vector Differences for Lexical Relation Learning <i>Vylomova, Rimell, Cohn, and Baldwin</i>	Target-Side Context for Discriminative Models in Statistical Machine Translation <i>Tamchyna, Fraser, Bojar, and Junczys-Dowmunt</i>	Situation entity types: automatic classification of clause-level aspect <i>Friedrich, Palmer, and Pinkal</i>	Generating Natural Questions About an Image <i>Mostafazadeh, Misra, Devlin, Mitchell, He, and Vanderwende</i>	Optimizing an Approximation of ROUGE - a Problem-Reduction Approach to Extractive Multi-Document Summarization <i>Peyrard and Eckle-Kohler</i>	User Modeling in Language Learning with Macaronic Texts <i>Renduchintala, Knowles, Koehn, and Eisner</i>
Cross-domain Text Classification with Multiple Domains and Disparate Label Sets <i>Bhatt, Sinha, and Roy</i>		Neural Machine Translation of Rare Words with Subword Units <i>Sennrich, Haddow, and Birch</i>	Online Adaptor Grammars with Hybrid inference <i>Pavlick and Tetreault</i>	Physical Causality of Action Verbs in Grounded Language Understanding <i>Gao, Doering, Yang, and Chai</i>	Online Adaptor Grammars with Hybrid inference <i>Xu, Napoles, Pavlick, Chen, and Callison-Burch</i>	On the Similarities Between Native, Non-native and Translated Texts <i>Rabinovich, Nisioi, Ordan, and Wintner</i>

03:30

03:50

04:10

04:30

04:50

Parallel Session 6

6A: Machine learning

Plaza Ballroom A & B

Chair:

Learning Structured Predictors from Bandit Feedback for Interactive NLP

Artem Sokolov, Julia Kreutzer, Christopher Lo, and Stefan Riezler

15:30–15:50

Structured prediction from bandit feedback describes a learning scenario where instead of having access to a gold standard structure, a learner only receives partial feedback in form of the loss value of a predicted structure. We present new learning objectives and algorithms for this interactive scenario, focusing on convergence speed and ease of elicitability of feedback. We present supervised-to-bandit simulation experiments for several NLP tasks (machine translation, sequence labeling, text classification), showing that bandit learning from relative preferences eases feedback strength and yields improved empirical convergence.

Online Adaptor Grammars with Hybrid inference

Silvana Hartmann, Judith Eckle-Kohler, and Iryna Gurevych

15:50–16:10

Deep Reinforcement Learning with a Natural Language Action Space

Ji He, Jianshu Chen, Xiaodong He, Jianfeng Gao, Lihong Li, Li Deng, and Mari Ostendorf
16:10–16:30

This paper introduces a novel architecture for reinforcement learning with deep neural networks designed to handle state and action spaces characterized by natural language, as found in text-based games. Termed a deep reinforcement relevance network (DRRN), the architecture represents action and state spaces with separate embedding vectors, which are combined with an interaction function to approximate the Q-function in reinforcement learning. We evaluate the DRRN on two popular text games, showing superior performance over other deep Q-learning architectures. Experiments with paraphrased action descriptions show that the model is extracting meaning rather than simply memorizing strings of text.

Incorporating Copying Mechanism in Sequence-to-Sequence Learning

Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. Li

16:30–16:50

We address an important problem in sequence-to-sequence (Seq2Seq) learning referred to as copying, in which certain segments in the input sequence are selectively replicated in the output sequence. A similar phenomenon is observable in human language communication. For example, humans tend to repeat entity names or even long phrases in conversation. The challenge with regard to copying in Seq2Seq is that new machinery is needed to decide when to perform the operation. In this paper, we incorporate copying into neural network-based Seq2Seq learning and propose a new model called CopyNet with encoder-decoder structure. CopyNet can nicely integrate the regular way of word generation in the decoder with the new copying mechanism which can choose sub-sequences in the input sequence and put them at proper places in the output sequence. Our empirical study on both synthetic data sets and real world data sets demonstrates the efficacy of CopyNet. For example, CopyNet can outperform regular RNN-based model with remarkable margins on text summarization tasks.

Cross-domain Text Classification with Multiple Domains and Disparate Label Sets

Himanshu Sharad Bhatt, Manjira Sinha, and Shourya Roy

16:50–17:10

Advances in transfer learning have let go the limitations of traditional supervised machine learning algorithms for being dependent on annotated training data for training new models for every new domains. Several real world applications encounter scenarios where models need to transfer/adapt across domains when the label sets vary both in terms of count of labels as well as their connotations. This paper presents first-of-its-kind transfer learning algorithm for cross-domain classification with multiple source domains and disparate label sets. It starts with identifying transferable knowledge from across multiple domains that can be useful for learning the target domain task. This knowledge in the form of selective labeled instances from different domains is congregated to form an auxiliary training set which is used for the target domain task. Experimental results validate the efficacy of the proposed algorithm against strong baselines on a real world social media and the \$20\$ Newsgroups datasets.

6B: Word vectors II

Plaza Ballroom D & E

Chair:

Online Adaptor Grammars with Hybrid inference*Sanjeev Arora, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski*

15:30–15:50

Semantic word embeddings represent the meaning of a word via a vector, and are created by diverse methods. Many use nonlinear operations on co-occurrence statistics, and have hand-tuned hyperparameters and reweightings. This paper gives a new generative model, a dynamic version of the log-linear topic model of Mnih and Hinton (2007). The methodological novelty is to use the prior to compute closed form expressions for word statistics. This provides a theoretical explanation for nonlinear models like PMI, word2vec, and GloVe, as well as some hyperparameter choices. It also helps explain why low-dimensional semantic embeddings contain linear algebraic structure that allows solution of word analogies, as shown by Mikolov et al. (2013a) and many subsequent papers. Experimental support is provided for the generative model assumptions, the most important of which is that latent word vectors are fairly uniformly dispersed in space.

Morphological Smoothing and Extrapolation of Word Embeddings*Ryan Cotterell, Hinrich Schütze, and Jason Eisner*

15:50–16:10

Languages with rich inflectional morphology exhibit lexical data sparsity, since the word used to express a given concept will vary with the syntactic context. For instance, each count noun in Czech has 12 forms (where English uses only singular and plural). Even in large corpora, we are unlikely to observe all inflections of a given lemma. This reduces the vocabulary coverage of methods that induce continuous representations for words from distributional corpus information. We solve this problem by exploiting existing morphological resources that can enumerate a word's component morphemes. We present a latent-variable Gaussian graphical model that allows us to extrapolate continuous representations for words not observed in the training corpus, as well as smoothing the representations provided for the observed words. The latent variables represent embeddings of morphemes, which combine to create embeddings of words. Over five languages and training sizes, our model improves the embeddings for words, when evaluated on an analogy task, skip-gram predictive accuracy and word similarity.

Cross-lingual Models of Word Embeddings: An Empirical Comparison*Shyam Upadhyay, Manaal Faruqui, Chris Dyer, and Dan Roth*

16:10–16:30

Despite interest in using cross-lingual knowledge to learn word embeddings for various tasks, a systematic comparison of the possible approaches is lacking in the literature. We perform an extensive evaluation of four popular approaches of inducing cross-lingual embeddings, each requiring a different form of supervision, on four typographically different language pairs. Our evaluation setup spans four different tasks, including intrinsic evaluation on mono-lingual and cross-lingual similarity, and extrinsic evaluation on downstream semantic and syntactic applications. We show that models which require expensive cross-lingual knowledge almost always perform better, but cheaply supervised models often prove competitive on certain tasks.

Take and Took, Gaggle and Goose, Book and Read: Evaluating the Utility of Vector Differences for Lexical Relation Learning*Ekaterina Vylomova, Laura Rimell, Trevor Cohn, and Timothy Baldwin*

16:30–16:50

Recent work has shown that simple vector subtraction over word embeddings is surprisingly effective at capturing different lexical relations, despite lacking explicit supervision. Prior work has evaluated this intriguing result using a word analogy prediction formulation and hand-selected relations, but the generality of the finding over a broader range of lexical relation types and different learning settings has not been evaluated. In this paper, we carry out such an evaluation in two learning settings: (1) spectral clustering to induce word relations, and (2) supervised learning to classify vector differences into relation types. We find that word embeddings capture a surprising amount of information, and that, under suitable supervised training, vector subtraction generalises well to a broad range of relations, including over unseen lexical items.

6C: Machine translation III

Plaza Ballroom F

*Chair:***Minimum Risk Training for Neural Machine Translation***Shiqi Shen, Yong Cheng, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu* 15:30–15:50

We propose minimum risk training for end-to-end neural machine translation. Unlike conventional maximum likelihood estimation, minimum risk training is capable of optimizing model parameters directly with respect to arbitrary evaluation metrics, which are not necessarily differentiable. Experiments show that our approach achieves significant improvements over maximum likelihood estimation on a state-of-the-art neural machine translation system across various languages pairs. Transparent to architectures, our approach can be applied to more neural networks and potentially benefit more NLP tasks.

Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations*Hoang Cuong, Khalil Sima'an, and Ivan Titov*

15:50–16:10

Existing work on domain adaptation for statistical machine translation has consistently assumed access to a small sample from the test distribution (target domain) at training time. In practice, however, the target domain may not be known at training time or it may change to match user needs. In such situations, it is natural to push the system to make safer choices, giving higher preference to domain invariant translations, which work well across domains, over risky domain-specific alternatives. We encode this intuition by (1) inducing latent subdomains from the training data only; (2) introducing features which measure how specialized phrases are to individual induced sub-domains; (3) estimating feature weights on out-of-domain data (rather than on the target domain). We conduct experiments on three language pairs and a number of different domains. We observe consistent improvements over a baseline which does not explicitly reward domain invariance.

A Character-level Decoder without Explicit Segmentation for Neural Machine Translation*Junyoung Chung, Kyunghyun Cho, and Yoshua Bengio*

16:10–16:30

The existing machine translation systems, whether phrase-based or neural, have relied almost exclusively on word-level modelling with explicit segmentation. In this paper, we ask a fundamental question: can neural machine translation generate a character sequence without any explicit segmentation? To answer this question, we evaluate an attention-based encoder–decoder with a subword-level encoder and a character-level decoder on four language pairs—En-Cs, En-De, En-Ru and En-Fi—using the parallel corpora from WMT’15. Our experiments show that the models with a character-level decoder outperform the ones with a subword-level decoder on all of the four language pairs. Furthermore, the ensembles of neural models with a character-level decoder outperform the state-of-the-art non-neural machine translation systems on En-Cs, En-De and En-Fi and perform comparably on En-Ru.

Target-Side Context for Discriminative Models in Statistical Machine Translation*Aleš Tamchyna, Alexander Fraser, Ondřej Bojar, and Marcin Junczys-Dowmunt* 16:30–16:50

Discriminative translation models utilizing source context have been shown to help statistical machine translation performance. We propose a novel extension of this work using target context information. Surprisingly, we show that this model can be efficiently integrated directly in the decoding process. Our approach scales to large training data sizes and results in consistent improvements in translation quality. We also provide an analysis comparing the strengths of the baseline source-context model with our extended source-context and target-context model and we show that our extension allows us to better capture morphological coherence. Our work is freely available as part of Moses.

Neural Machine Translation of Rare Words with Subword Units*Rico Sennrich, Barry Haddow, and Alexandra Birch*

16:50–17:10

Neural machine translation (NMT) models typically operate with a fixed vocabulary, but translation is an open-vocabulary problem. Previous work addresses the translation of out-of-vocabulary words by back-*off* to a dictionary. In this paper, we introduce a simpler and more effective approach, making the NMT model capable of open-vocabulary translation by encoding rare and unknown words as sequences of subword units. This is based on the intuition that various word classes are translatable via smaller units than words, for instance names (via character copying or transliteration), compounds (via compositional translation), and cognates and loanwords (via phonological and morphological transformations). We discuss the suitability of different word segmentation techniques, including simple character n-gram models and a segmentation based on the byte pair encoding compression algorithm, and empirically show that subword models improve over a back-off dictionary baseline for the WMT 15 translation tasks English-German and English-Russian by up to 1.1 and 1.3 BLEU, respectively.

6D: Discourse*Chair:***Implicit Discourse Relation Detection via a Deep Architecture with Gated Relevance Network***Jifan Chen, Qi Zhang, Pengfei Liu, Xipeng Qiu, and Xuanjing Huang*

15:30–15:50

Word pairs, which are one of the most easily accessible features between two text segments, have been proven to be very useful for detecting the discourse relations held between text segments. However, because of the data sparsity problem, the performance achieved by using word pair features is limited. In this paper, in order to overcome the data sparsity problem, we propose the use of word embeddings to replace the original words. Moreover, we adopt a gated relevance network to capture the semantic interaction between word pairs, and then aggregate those semantic interactions using a pooling layer to select the most informative interactions. Experimental results on Penn Discourse Tree Bank show that the proposed method without using manually designed features can achieve better performance on recognizing the discourse level relations in all of the relations.

Model Architectures for Quotation Detection*Christian Scheible, Roman Klinder, and Sebastian Padó*

15:50–16:10

Quotation detection is the task of locating spans of quoted speech in text. The state of the art treats this problem as a sequence labeling task and employs linear-chain conditional random fields. We question the efficacy of this choice: The Markov assumption in the model prohibits it from making joint decisions about the begin, end, and internal context of a quotation. We perform an extensive analysis with two new model architectures. We find that (a), simple boundary classification combined with a greedy prediction strategy is competitive with the state of the art; (b), a semi-Markov model significantly outperforms all others, by relaxing the Markov assumption.

Speech Act Modeling of Written Asynchronous Conversations with Task-Specific Embeddings and Conditional Structured Models*Shafiq Joty and Enamul Hoque*

16:10–16:30

This paper addresses the problem of speech act recognition in written asynchronous conversations (e.g., fora, emails). We propose a class of conditional structured models defined over arbitrary graph structures to capture the conversational dependencies between sentences. Our models use sentence representations encoded by a long short term memory (LSTM) recurrent neural model. Empirical evaluation shows the effectiveness of our approach over existing ones: LSTMs provide better task-specific representations, and the global joint model improves over local models.

Situation entity types: automatic classification of clause-level aspect*Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal*

16:30–16:50

This paper describes the first robust approach to automatically labeling clauses with their situation entity type (Smith 2003), capturing aspectual phenomena at the clause level which are relevant for interpreting both semantics at the clause level and discourse structure. Previous work on this task used a small data set from a limited domain, and relied mainly on words as features, an approach which is impractical in larger settings. We provide a new corpus of texts from 13 genres (40,000 clauses) annotated with situation entity types. We show that our sequence labeling approach using distributional information in the form of Brown clusters, as well as syntactic-semantic features targeted to the task, is robust across genres, reaching accuracies of up to 76%.

Online Adaptor Grammars with Hybrid inference*Ellie Pavlick and Joel Tetreault*

16:50–17:10

This paper presents an empirical study of linguistic formality. We perform an analysis of humans' perceptions of formality in four different genres. These findings are used to develop a statistical model for predicting formality, which is evaluated under different feature settings and genres. We apply our model to an investigation of formality in online discussion forums, and present findings consistent with theories of formality and linguistic coordination.

6E: Language and vision

Chair:

Learning Prototypical Event Structure from Photo Albums*Antoine Bosselut, Jianfu Chen, David Warren, Hannaneh Hajishirzi, and Yejin Choi* 15:30–15:50

Activities and events in our lives are structural, be it a vacation, a camping trip, or a wedding. While individual details vary, there are characteristic patterns that are specific to each of these scenarios. For example, a wedding typically consists of a sequence of events such as walking down the aisle, exchanging vows, and dancing. In this paper, we present a data-driven approach to learning event knowledge from a large collection of photo albums. We formulate the task as constrained optimization to induce the prototypical temporal structure of an event, integrating both visual and textual cues. Comprehensive evaluation demonstrates that it is possible to learn multimodal knowledge of event structure from noisy web content.

Cross-Lingual Image Caption Generation*Takashi Miyazaki and Nobuyuki Shimizu*

15:50–16:10

A fundamental problem in artificial intelligence is automatically generating a natural language description of an image. This task involves both computer vision and natural language processing and is called “image caption generation.” Research on image caption generation has typically focused on taking an image and generating a caption in English as existing image caption corpora are mostly in English. The lack of corpora in languages other than English is an issue, especially for morphologically rich languages such as Japanese. There is thus a need for corpora sufficiently large for image captioning in other languages. We have developed a Japanese version of the MS COCO caption dataset and a generative model based on a deep recurrent architecture that takes in an image and uses this Japanese version of the dataset to generate a caption in Japanese. As the Japanese portion of the corpus is small, our model was designed to transfer the knowledge representation obtained from the English portion into the Japanese portion. Experiments showed that the resulting bilingual comparable corpus has better performance than a monolingual corpus, indicating that image understanding using a resource-rich language benefits a resource-poor language.

Learning Concept Taxonomies from Multi-modal Data*Hao Zhang, Zhiteng Hu, Yuntian Deng, Mrinmaya Sachan, Zhicheng Yan, and Eric Xing* 16:10–16:30

We study the problem of automatically building hypernym taxonomies from textual and visual data. Previous works in taxonomy induction generally ignore the increasingly prominent visual data, which encode important perceptual semantics. Instead, we propose a probabilistic model for taxonomy induction by jointly leveraging text and images. To avoid hand-crafted feature engineering, we design end-to-end features based on distributed representations of images and words. The model is discriminatively trained given a small set of existing ontologies and is capable of building full taxonomies from scratch for a collection of unseen conceptual label items with associated images. We evaluate our model and features on the WordNet hierarchies, where our system outperforms previous approaches by a large gap.

Generating Natural Questions About an Image*Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Margaret Mitchell, Xiaodong He, and Lucy Vanderwende*

16:30–16:50

There has been an explosion of work in the vision & language community during the past few years from image captioning to video transcription, and answering questions about images. These tasks have focused on literal descriptions of the image. To move beyond the literal, we choose to explore how questions about an image are often directed at commonsense inference and the abstract events evoked by objects in the image. In this paper, we introduce the novel task of Visual Question Generation, where the system is tasked with asking a natural and engaging question when shown an image. We provide three datasets which cover a variety of images from object-centric to event-centric, with considerably more abstract training data than provided to state-of-the-art captioning systems thus far. We train and test several generative and retrieval models to tackle the task of VQG. Evaluation results show that while such models ask reasonable questions for a variety of images, there is still a wide gap with human performance which motivates further work on connecting images with commonsense knowledge and pragmatics. Our proposed task offers a new challenge to the community which we hope furthers interest in exploring deeper connections between vision & language.

Physical Causality of Action Verbs in Grounded Language Understanding*Qiaozhi Gao, Malcolm Doering, Shaohua Yang, and Joyce Chai*

16:50–17:10

Linguistics studies have shown that action verbs often denote some Change of State (CoS) as the result of an action. However, the causality of action verbs and its potential connection with the physical world has not been systematically explored. To address this limitation, this paper presents a study on physical causality of action verbs and their implied changes in the physical world. We first conducted a crowd-sourcing experiment and identified eighteen categories of physical causality for action verbs. For a subset of these categories, we then defined a set of detectors that detect the corresponding change from visual perception of the physical environment. We further incorporated physical causality modeling and state detection in grounded language understanding. Our empirical studies have demonstrated the effectiveness of causality modeling in grounding language to perception.

6F: Summarization*Chair:***Optimizing an Approximation of ROUGE - a Problem-Reduction Approach to Extractive Multi-Document Summarization***Maxime Peyrard and Judith Eckle-Kohler*

16:30–16:50

This paper presents a problem-reduction approach to extractive multi-document summarization: we propose a reduction to the problem of scoring individual sentences with their ROUGE scores based on supervised learning. For the summarization, we solve an optimization problem where the ROUGE score of the selected summary sentences is maximized. To this end, we derive an approximation of the ROUGE-N score of a set of sentences, and define a principled discrete optimization problem for sentence selection. Mathematical and empirical evidence suggests that the sentence selection step is solved almost exactly, thus reducing the problem to the sentence scoring task. We perform a detailed experimental evaluation on two DUC datasets to demonstrate the validity of our approach.

Online Adaptor Grammars with Hybrid inference*Wei Xu, Courtney Napolis, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch* 16:50–17:10

Most recent sentence simplification systems use basic machine translation models to learn lexical and syntactic paraphrases from a manually simplified parallel corpus. These methods are limited by the quality and quantity of manually simplified corpora, which are expensive to build. In this paper, we conduct an in-depth adaptation of statistical machine translation to perform text simplification, taking advantage of large-scale paraphrases learned from bilingual texts and a small amount of manual simplifications with multiple references. Our work is the first to design automatic metrics that are effective for tuning and evaluating simplification systems, which will facilitate iterative development for this task.

6G: Learner language

Chair:

Phrase Structure Annotation and Parsing for Learner English

Ryo Nagata and Keisuke Sakaguchi

15:30–15:50

There has been almost no work on phrase structure annotation and parsing specially designed for learner English despite the fact that they are useful for representing the structural characteristics of learners. To address this problem, in this paper, we first propose a phrase structure annotation scheme for learner English and annotate two different learner corpora using it. Second, we show their usefulness, reporting on (a) inter-annotator agreement rate, (b) characteristic CFG rules in the corpora, and (c) parsing performance on them. In addition, we explore methods to improve phrase structure parsing for learner English (achieving an F-measure of 0.878). Finally, we release the full annotation guidelines, the annotated data, and the improved parser model for learner English to the public.

A Trainable Spaced Repetition Model for Language Learning

Burr Settles and Brendan Meeder

15:50–16:10

We present half-life regression (HLR), a novel model for spaced repetition practice with applications to second language acquisition. HLR combines psycholinguistic theory with modern machine learning techniques, indirectly estimating the “half-life” of a word or concept in a student’s long-term memory. We use data from Duolingo — a popular online language learning application — to fit HLR models, reducing error by 45%+ compared to several baselines at predicting student recall rates. HLR model weights also shed light on which linguistic concepts are systematically challenging for second language learners. Finally, HLR was able to improve Duolingo daily student engagement by 12% in an operational user study.

Reassessing the Goals of Grammatical Error Correction: Fluency Instead of Grammaticality

Keisuke Sakaguchi, Courtney Napoles, Matt Post, and Joel Tetreault

16:10–16:30

The field of grammatical error correction (GEC) has grown substantially in recent years, with research directed at both evaluation metrics and improved system performance against those metrics. One unvisited assumption, however, is the reliance of GEC evaluation on error-coded corpora, which contain specific labeled corrections. We examine current practices and show that GEC’s reliance on such corpora unnaturally constrains annotation and automatic evaluation, resulting in (a) sentences that do not sound acceptable to native speakers and (b) system rankings that do not correlate with human judgments. In light of this, we propose an alternate approach that jettisons costly error coding in favor of unannotated, whole-sentence rewrites. We compare the performance of existing metrics over different gold-standard annotations, and show that automatic evaluation with our new annotation scheme has very strong correlation with expert rankings ($= 0.82$). As a result, we advocate for a fundamental and necessary shift in the goal of GEC, from correcting small, labeled error types, to producing text that has native fluency.

User Modeling in Language Learning with Macaronic Texts

Adithya Renduchintala, Rebecca Knowles, Philipp Koehn, and Jason Eisner

16:30–16:50

Foreign language learners can acquire new vocabulary by using cognate and context clues when reading. We devise an experimental framework to measure such incidental comprehension using Amazon Mechanical Turk (MTurk). With the data collected, we train a graphical model to jointly predict which words users will be able to comprehend, using features such as edit distance (cognate clues) and pointwise mutual information (context clues). The results show that our model’s predictions correlate with learner understanding, enabling the construction of comprehensible text for personalized foreign language education.

On the Similarities Between Native, Non-native and Translated Texts

Ella Rabinovich, Sergiu Nisioi, Noam Ordan, and Shuly Wintner

16:50–17:10

We present a computational analysis of three varieties of language: native, advanced non-native, and translation. Our goal is to investigate the similarities and differences between non-native language productions and translations, contrasting both with native language. Using a collection of computational methods we establish three main results: (1) the three types of texts are easily distinguishable; (2) non-native language and translations are closer to each other than each of them is to native language; and (3) some of these characteristics are source- and native-language-dependent, while others are not, reflecting, perhaps, unified principles that similarly affect translations and non-native language.

Main Conference: Wednesday, June 3

Overview

9:00–9:40	President's talk	'sLoc			
9:40–10:10	Coffee break				
10:10–11:50	SESSION [('Wednesday', 'August 10', '2016')/Session] 7A Outstanding papers I				
10:10–11:50	SESSION [('Wednesday', 'August 10', '2016')/Session] 7B Outstanding papers II				
11:50–1:30	Lunch break				
1:30–3:00	ACL business meeting (open to all)				
3:00–3:30	Coffee break				
	Session 8				
3:30–4:50	Question answering II (short papers)	Word vectors III (short papers)	Topics and discourse (short papers)	Syntax and morphology (short papers)	Potpourri II (short papers)
5:00–6:30	Awards session				
6:30–7:00	Closing session				

Session 8 Overview – Wednesday, June 3, 2015

	Track A <i>Question answering II (short papers)</i>	Track B <i>Word vectors III (short papers)</i>	Track C <i>Topics and discourse (short papers)</i>	Track D <i>Syntax and morphology (short papers)</i>	Track E <i>Potpourri II (short papers)</i>
03:30	Plaza Ballroom A & Plaza Ballroom D & Plaza Ballroom F	Cross-Lingual Word Representations via Spectral Graph Embeddings <i>Oshikiri, Fukui, and Shimodaira</i>	Claim Synthesis via Predicate Recycling <i>Bilu and Slonim</i>	Single-Model Encoder-Decoder with Explicit Morphological Representation for Reinflection <i>Kann and Schütze</i>	Metrics for Evaluation of Word-level Machine Translation Quality Estimation <i>Logacheva, Lukasik, and Specia</i>
03:46	Science Question Answering using Instructional Materials <i>Sachan, Dubey, and Xing</i>	Semantics-Driven Recognition of Collocations Using Word Embeddings <i>Rodríguez-Fernández, Espinosa Anke, Carlini, and Wanner</i>	Modelling the Interpretation of Discourse Connectives by Bayesian Pragmatics <i>Yung, Duh, Komura, and Matsumoto</i>	Joint part-of-speech and dependency projection from multiple sources <i>Johannsen, Agić, and Søgaard</i>	The Social Impact of Natural Language Processing <i>Hovy and Spruit</i>
04:02	Specifying and Annotating Reduced Argument Span Via QA-SRL <i>Stanovsky, Dagan, and Adler</i>	Incorporating Relational Knowledge into Word Representations using Subspace Regularization <i>Kumar and Araki</i>	Nonparametric Spherical Topic Modeling with Word Embeddings <i>Batmanghelich, Saeedi, Narasimhan, and Gershman</i>	Dependency-based Gated Recursive Neural Network for Chinese Word Segmentation <i>Xu and Sun</i>	Using Sequence Similarity Networks to Identify Partial Cognates in Multilingual Wordlists <i>List, Lopez, and Baptiste</i>
04:18	Improving Argument Overlap for Proposition-Based Summarisation <i>Fang and Teufel</i>	Word Embedding Calculus in Meaningful Ultradense Subspaces <i>Rothe and Schütze</i>	A Novel Measure for Coherence in Statistical Topic Models <i>Morstatter and Liu</i>	Deep Neural Networks for Syntactic Parsing of Morphologically Rich Languages <i>Legrand and Collerbert</i>	
04:34	Machine Comprehension using Rich Semantic Representations <i>Sachan and Xing</i>	Is “Universal Syntax” Universally Useful for Learning Distributed Word Representations? <i>Vulić and Korhonen</i>	Coarse-grained Argumentation Features for Scoring Persuasive Essays <i>Ghosh, Khanam, Han, and Muresan</i>	Weakly Supervised Part-of-speech Tagging Using Eye-tracking Data <i>Barrett, Bingel, Keller, and Søgaard</i>	

Parallel Session 8

8A: Question answering II (short papers)

Plaza Ballroom A & B

Chair:

Machine Translation Evaluation Meets Community Question Answering

Francisco Guzmán, Lluís Márquez, and Preslav Nakov

15:30–15:46

We explore the applicability of machine translation evaluation (MTE) methods to a very different problem: answer ranking in community Question Answering. In particular, we adopt a pairwise neural network (NN) architecture, which incorporates MTE features, as well as rich syntactic and semantic embeddings, and which efficiently models complex non-linear interactions. The evaluation results show state-of-the-art performance, with sizeable contribution from both the MTE features and from the pairwise NN architecture.

Science Question Answering using Instructional Materials

Mrinmaya Sachan, Kumar Dubey, and Eric Xing

15:46–16:02

We provide a solution for elementary science test using instructional materials. We posit that there is a hidden structure that explains the correctness of an answer given the question and instructional materials and present a unified max-margin framework that learns to find these hidden structures (given a corpus of question-answer pairs and instructional materials), and uses what it learns to answer novel elementary science questions. Our evaluation shows that our framework outperforms several strong baselines.

Specifying and Annotating Reduced Argument Span Via QA-SRL

Gabriel Stanovsky, Ido Dagan, and Meni Adler

16:02–16:18

Prominent semantic annotations take an inclusive approach to argument span annotation, marking arguments as full constituency subtrees. Some works, however, showed that identifying a reduced argument span can be beneficial for various semantic tasks. While certain practical methods do extract reduced argument spans, such as in OpenIE, these solutions are often ad-hoc and system-dependent, with no commonly accepted standards. In this paper we propose a generic argument reduction criterion, along with an annotation procedure, and show that it can be consistently and intuitively annotated using the recent QA-SRL paradigm.

Improving Argument Overlap for Proposition-Based Summarisation

Yimai Fang and Simone Teufel

16:18–16:34

We present improvements to our incremental proposition-based summariser, which is inspired by Kintsch and van Dijk's (1978) text comprehension model. Argument overlap is a central concept in this summariser. Our new model replaces the old overlap method based on distributional similarity with one based on lexical chains. We evaluate on a new corpus of 124 summaries of educational texts, and show that our new system outperforms the old method and several state-of-the-art non-proposition-based summarisers. The experiment also verifies that the incremental nature of memory cycles is beneficial in itself, by comparing it to a non-incremental algorithm using the same underlying information.

Machine Comprehension using Rich Semantic Representations

Mrinmaya Sachan and Eric Xing

16:34–16:50

Machine comprehension tests the system's ability to understand a piece of text through a reading comprehension task. For this task, we propose an approach using the Abstract Meaning Representation (AMR) formalism. We construct meaning representation graphs for the given text and for each question-answer pair by merging the AMRs of comprising sentences using cross-sentential phenomena such as coreference and rhetorical structures. Then, we reduce machine comprehension to a graph containment problem. We posit that there is a latent mapping of the question-answer meaning representation graph onto the text meaning representation graph that explains the answer. We present a unified max-margin framework that learns to find this mapping (given a corpus of texts and question-answer pairs), and uses what it learns to answer questions on novel texts. We show that this approach leads to state of the art results on the task.

8B: Word vectors III (short papers)

Plaza Ballroom D & E

Chair:

Cross-Lingual Word Representations via Spectral Graph Embeddings

Takamasa Oshikiri, Kazuki Fukui, and Hideyoshi Shimodaira

15:30–15:46

Cross-lingual word embeddings are used for cross-lingual information retrieval or domain adaptations. In this paper, we extend Eigenwords, spectral monolingual word embeddings based on canonical correlation analysis (CCA), to cross-lingual settings with sentence-alignment. For incorporating cross-lingual information, CCA is replaced with its generalization based on the spectral graph embeddings. The proposed method, which we refer to as Cross-Lingual Eigenwords (CL-Eigenwords), is fast and scalable for computing distributed representations of words via eigenvalue decomposition. Numerical experiments of English-Spanish word translation tasks show that CL-Eigenwords is competitive with state-of-the-art cross-lingual word embedding methods.

Semantics-Driven Recognition of Collocations Using Word Embeddings

Sara Rodríguez-Fernández, Luis Espinosa Anke, Roberto Carlini, and Leo Wanner 15:46–16:02

L2 learners often produce “ungrammatical” word combinations such as, e.g., *give a suggestion or *make a walk. This is because of the “collocationality” of one of their items (the base) that limits the acceptance of collocates to express a specific meaning (‘perform’ above). We propose an algorithm that delivers, for a given base and the intended meaning of a collocate, the actual collocate lexeme(s) (make / take above). The algorithm exploits the linear mapping between bases and collocates from examples and generates a collocation transformation matrix which is then applied to novel unseen cases. The evaluation shows a promising line of research in collocation discovery.

Incorporating Relational Knowledge into Word Representations using Subspace Regularization

Abhishek Kumar and Jun Araki

16:02–16:18

Incorporating lexical knowledge from semantic resources (e.g., WordNet) has been shown to improve the quality of distributed word representations. This knowledge often comes in the form of relational triplets (x, r, y) where words x and y are connected by a relation type r . Existing methods either ignore the relation types, essentially treating the word pairs as generic related words, or employ rather restrictive assumptions to model the relational knowledge. We propose a novel approach to model relational knowledge based on low-rank subspace regularization, and conduct experiments on standard tasks to evaluate its effectiveness.

Word Embedding Calculus in Meaningful Ultradense Subspaces

Sascha Rothe and Hinrich Schütze

16:18–16:34

We decompose a standard embedding space into interpretable orthogonal subspaces, including a one-dimensional polarity subspace as well as concreteness, frequency and part-of-speech (POS) subspaces. We introduce a new calculus for subspaces that supports operations like “ $-1 \text{ X hate} = \text{love}$ ” and “give me a neutral word for greasy” (i.e., oleaginous). This calculus extends analogy computations like “king - man + woman = queen”. For the tasks of Antonym Classification and POS Tagging our method outperforms the state of the art. We create new test sets for Morphological Analogies and for the task of Polarity Spectrum Creation.

Is “Universal Syntax” Universally Useful for Learning Distributed Word Representations?

Ivan Vulić and Anna Korhonen

16:34–16:50

Recent comparative studies have demonstrated the usefulness of dependency-based contexts (DEPS) for learning distributed word representations for similarity tasks. In English, DEPS tend to perform better than the more common, less informed bag-of-words contexts (BOW). In this paper, we present the first cross-linguistic comparison of different context types for three different languages. DEPS are extracted from “universal parses” without any language-specific optimization. Our results suggest that the universal DEPS (UDEPS) are useful for detecting functional similarity (e.g., verb similarity, solving syntactic analogies) among languages, but their advantage over BOW is not as prominent as previously reported on English. We also show that simple “post-parsing” filtering of useful UDEPS contexts leads to consistent improvements across languages.

8C: Topics and discourse (short papers)

Plaza Ballroom F

Chair:

Claim Synthesis via Predicate Recycling*Yonatan Bilu and Noam Slonim*

15:30–15:46

Computational Argumentation has two main goals - the detection and analysis of arguments on the one hand, and the synthesis of arguments on the other. Much attention has been given to the former, but considerably less to the latter. A key component in synthesizing arguments is the synthesis of claims. One way to do so is by employing argumentation mining to detect claims within an appropriate corpus. In general, this appears to be a hard problem. Thus, it is interesting to explore if - for the sake of synthesis - there may be other ways to generate claims. Here we explore such a method: we extract the predicate of simple, manually-detected, claims, and attempt to generate novel claims from them. Surprisingly, this simple method yields fairly good results.

Modelling the Interpretation of Discourse Connectives by Bayesian Pragmatics*Frances Yung, Kevin Duh, Taku Komura, and Yuji Matsumoto*

15:46–16:02

We propose a framework to model human comprehension of discourse connectives. Following the Bayesian pragmatic paradigm, we advocate that discourse connectives are interpreted based on a simulation of the production process by the speaker, who, in turn, considers the ease of interpretation for the listener when choosing connectives. Evaluation against the sense annotation of the Penn Discourse Treebank confirms the superiority of the model over literal comprehension. A further experiment demonstrates that the proposed model also improves automatic discourse parsing.

Nonparametric Spherical Topic Modeling with Word Embeddings*Kayhan Batmanghelich, Ardavan Saeedi, Karthik Narasimhan, and Sam Gershman* 16:02–16:18

Traditional topic models do not account for semantic regularities in language. Recent distributional representations of words exhibit semantic consistency over directional metrics such as cosine similarity. However, neither categorical nor Gaussian observational distributions used in existing topic models are appropriate to leverage such correlations. In this paper, we propose to use the von Mises-Fisher distribution to model the density of words over a unit sphere. Such a representation is well-suited for directional data. We use a Hierarchical Dirichlet Process for our base topic model and propose an efficient inference algorithm based on Stochastic Variational Inference. This model enables us to naturally exploit the semantic structures of word embeddings while flexibly discovering the number of topics. Experiments demonstrate that our method outperforms three competitive approaches in terms of topic coherence on two different text corpora while offering efficient inference.

A Novel Measure for Coherence in Statistical Topic Models*Fred Morstatter and Huan Liu*

16:18–16:34

Big data presents new challenges for understanding large text corpora. Topic modeling algorithms help understand the underlying patterns, or “topics”, in data. Researchers author often read these topics in order to gain an understanding of the underlying corpus. It is important to evaluate the interpretability of these automatically generated topics. Methods have previously been designed to use crowdsourcing platforms to measure interpretability. In this paper, we demonstrate the necessity of a key concept, coherence, when assessing the topics and propose an effective method for its measurement. We show that the proposed measure of coherence captures a different aspect of the topics than existing measures. We further study the automation of these topic measures for scalability and reproducibility, showing that these measures can be automated.

Coarse-grained Argumentation Features for Scoring Persuasive Essays*Debanjan Ghosh, Aquila Khanam, Yubo Han, and Smaranda Muresan*

16:34–16:50

Scoring the quality of persuasive essays is an important goal of discourse analysis, addressed most recently with high-level persuasion-related features such as thesis clarity, or opinions and their targets. We investigate whether argumentation features derived from a coarse-grained argumentative structure of essays can help predict essays scores. We introduce a set of argumentation features related to argument components (e.g., the number of claims and premises), argument relations (e.g., the number of supported claims) and typology of argumentative structure (chains, trees). We show that these features are good predictors of human scores for TOEFL essays, both when the coarse-grained argumentative structure is manually annotated and automatically predicted.

8D: Syntax and morphology (short papers)

Chair:

Single-Model Encoder-Decoder with Explicit Morphological Representation for Reinflection

Katharina Kann and Hinrich Schütze

15:30–15:46

Morphological reinflection is the task of generating a target form given a source form, a source tag and a target tag. We propose a new way of modeling this task with neural encoder-decoder models. Our approach reduces the amount of required training data for this architecture and achieves state-of-the-art results, making encoder-decoder models applicable to morphological reinflection even for low-resource languages. We further present a new automatic correction method for the outputs based on edit trees.

Joint part-of-speech and dependency projection from multiple sources

Anders Johannsen, Željko Agić, and Anders Søgaard

15:46–16:02

Most previous work on annotation projection has been limited to a subset of Indo-European languages, using only a single source language, and projecting annotation for one task at a time. In contrast, we present an Integer Linear Programming (ILP) algorithm that simultaneously projects annotation for multiple tasks from multiple source languages, relying on parallel corpora available for hundreds of languages. When training POS taggers and dependency parsers on jointly projected POS tags and syntactic dependencies using our algorithm, we obtain better performance than a standard approach on 20/23 languages using one parallel corpus; and 18/27 languages using another.

Dependency-based Gated Recursive Neural Network for Chinese Word Segmentation

Jingjing Xu and Xu Sun

16:02–16:18

Recently, many neural network models have been applied to Chinese word segmentation. However, such models focus more on collecting local information while long distance dependencies are not well learned. To integrate local features with long distance dependencies, we propose a dependency-based gated recursive neural network. Local features are first collected by bi-directional long short term memory network, then combined and refined to long distance dependencies via gated recursive neural network. Experimental results show that our model is a competitive model for Chinese word segmentation.

Deep Neural Networks for Syntactic Parsing of Morphologically Rich Languages

Joël Legrand and Ronan Collobert

16:18–16:34

Morphologically rich languages (MRL) are languages in which much of the structural information is contained at the word-level, leading to high level word-form variation. Historically, syntactic parsing has been mainly tackled using generative models. These models assume input features to be conditionally independent, making difficult to incorporate arbitrary features. In this paper, we investigate the greedy discriminative parser described in (Legrand and Collobert, 2015), which relies on word embeddings, in the context of MRL. We propose to learn morphological embeddings and propagate morphological information through the tree using a recursive composition procedure. Experiments show that such embeddings can dramatically improve the average performance on different languages. Moreover, it yields state-of-the art performance for a majority of languages.

Weakly Supervised Part-of-speech Tagging Using Eye-tracking Data

Maria Barrett, Joachim Bingel, Frank Keller, and Anders Søgaard

16:34–16:50

For many of the world’s languages, there are no or very few linguistically annotated resources. On the other hand, raw text, and often also dictionaries, can be harvested from the web for many of these languages, and part-of-speech taggers can be trained with these resources. At the same time, previous research shows that eye-tracking data, which can be obtained without explicit annotation, contains clues to part-of-speech information. In this work, we bring these two ideas together and show that given raw text, a dictionary, and eye-tracking data obtained from naive participants reading text, we can train a weakly supervised PoS tagger using a second-order HMM with maximum entropy emissions. The best model use type-level aggregates of eye-tracking data and significantly outperforms a baseline that does not have access to eye-tracking data.

8E: Potpourri II (short papers)*Chair:***Metrics for Evaluation of Word-level Machine Translation Quality Estimation***Varvara Logacheva, Michal Lukasik, and Lucia Specia*

15:30–15:46

The aim of this paper is to investigate suitable evaluation strategies for the task of word-level quality estimation of machine translation. We suggest various metrics to replace F_1 -score for the “BAD” class, which is currently used as main metric. We compare the metrics’ performance on real system outputs and synthetically generated datasets and suggest a reliable alternative to the F_1 -BAD score — the multiplication of F_1 -scores for different classes. Other metrics have lower discriminative power and are biased by unfair labellings.

The Social Impact of Natural Language Processing*Dirk Hovy and Shannon L. Spruit*

15:46–16:02

Medical sciences have long since established an ethics code for experiments, to minimize the risk of harm to subjects. Natural language processing (NLP) used to involve mostly anonymous corpora, with the goal of enriching linguistic analysis, and was therefore unlikely to raise ethical concerns. As NLP becomes increasingly wide-spread and uses more data from social media, however, the situation has changed: the outcome of NLP experiments and applications can now have a direct effect on individual users’ lives. Until now, the discourse on this topic in the field has not followed the technological development, while public discourse was often focused on exaggerated dangers. This position paper tries to take back the initiative and start a discussion. We identify a number of social implications of NLP and discuss their ethical significance, as well as ways to address them.

Using Sequence Similarity Networks to Identify Partial Cognates in Multilingual Wordlists*Johann-Mattis List, Philippe Lopez, and Eric Baptiste*

16:02–16:18

Increasing amounts of digital data in historical linguistics necessitate the development of automatic methods for the detection of cognate words across languages. Recently developed methods work well on language families with moderate time depths, but they are not capable of identifying cognate morphemes in words which are only partially related. Partial cognacy, however, is a frequently recurring phenomenon, especially in language families with productive derivational morphology. This paper presents a pilot approach for partial cognate detection in which networks are used to represent similarities between word parts and cognate morphemes are identified with help of state-of-the-art algorithms for network partitioning. The approach is tested on a newly created benchmark dataset with data from three sub-branches of Sino-Tibetan and yields very promising results, outperforming all algorithms which are not sensible to partial cognacy.

]

6

*SEM

*SEM 2016: The 5th Joint Conference on Lexical and Computational Semantics

Organizers: Claire Gardent, Raffaella Bernardi, Ivan Titov, Phong Le, Kevin Duh, Diana McCarthy,
Roberto Navigli, Jonathan Ginzburg, Yoav Artzi, Yonatan Bisk, Annie Louis,
Michael Roth and Elena Cabrio

Venue: Governor's Square 14

Thursday, August 11, 2016

9:00–9:10 **Welcome**

9:10–10:00 **Invited Talk: Context and Non-compositional Phenomena in Language Understanding (Yoav Artzi)**

10:00–10:30 Quantificational features in distributional word representations
Tal Linzen, Emmanuel Dupoux, and Benjamin Spector

10:30–11:00 **Break**

Lexical semantics

11:00–11:30 Automatic Identification of Aspectual Classes across Verbal Readings
Ingrid Falk and Fabienne Martin

11:30–12:00 Metaphor as a Medium for Emotion: An Empirical Study
Saif Mohammad, Ekaterina Shutova, and Peter Turney

12:00–12:30 High-Fidelity Lexical Axiom Construction from Verb Glosses
Gene Kim and Lenhart Schubert

12:30–2:00 **Lunch break**

Semantic parsing and formal semantics

2:00–2:30 Implicit Semantic Roles in a Multilingual Setting
Jennifer Sikos, Yannick Versley, and Anette Frank

2:30–3:00 Driving inversion transduction grammar induction with semantic evaluation
Meriem Beloucif and Dekai Wu

3:00–3:30 Natural Solution to FraCaS Entailment Problems
Lasha Abzianidze

3:30–4:00 **Break**

Formal and linguistic semantics

4:00–4:30 How Factuality Determines Sentiment Inferences
Manfred Klenner and Simon Clematide

4:30–5:30 **Poster Session**

- Sense Embedding Learning for Word Sense Induction
Linfeng Song, Zhiguo Wang, Haitao Mi, and Daniel Gildea
- Improving Zero-Shot-Learning for German Particle Verbs by using Training-Space Restrictions and Local Scaling
Maximilian Köper, Sabine Schulte im Walde, Max Kisselw, and Sebastian Padó
- When Hyperparameters Help: Beneficial Parameter Combinations in Distributional Semantic Models
Alicia Krebs and Denis Paperno
- Leveraging VerbNet to build Corpus-Specific Verb Clusters
Daniel Peterson, Jordan Boyd-Graber, Martha Palmer, and Daisuke Kawahara
- Adding Context to Semantic Data-Driven Paraphrasing
Vered Shwartz and Ido Dagan
- So-Called Non-Subsective Adjectives
Ellie Pavlick and Chris Callison-Burch
- Linguistic Style Accommodation in Disagreements
Elise van der Pol, Sharon Gieske, and Raquel Fernandez
- Unsupervised Text Segmentation Using Semantic Relatedness Graphs
Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto

- Improving Text-to-Pictograph Translation Through Word Sense Disambiguation
Leen Sevens, Gilles Jacobs, Vincent Vandeghinste, Ineke Schuurman, and Frank Van Eynde
- Taking the best from the Crowd: Learning Question Passage Classification from Noisy Data
Azad Abad and Alessandro Moschitti
- Orthogonality regularizer for question answering
Chunyang Xiao, Guillaume Bouchard, Marc Dymetman, and Claire Gardent

Friday, August 12, 2016

- 9:10–10:00 **Invited talk: Top-down and bottom-up views on success in semantics (Alexander Koller)**
- 10:00–10:30 The Role of Modifier and Head Properties in Predicting the Compositionality of English and German Noun-Noun Compounds: A Vector-Space Perspective
Sabine Schulte im Walde, Anna Häfty, and Stefan Bott
- 10:30–11:00 **Break**
- Semantics for applications and distributional semantics**
- 11:00–11:30 Detecting Stance in Tweets And Analyzing its Interaction with Sentiment
Parinaz Sobhani, Saif Mohammad, and Svetlana Kiritchenko
- 11:30–12:00 A Study of Suggestions in Opinionated Texts and their Automatic Detection
Sapna Negi, Kartik Asooja, Shubham Mehrotra, and Paul Buitelaar
- 12:00–12:30 You and me... in a vector space: modelling individual speakers with distributional semantics
Aurélie Herbelot and Behrang QasemiZadeh
- 12:30–2:10 **Lunch break**
- 2:10–3:00 **Invited talk: Exploring for Concurrent Discourse Relations (Bonnie Webber)**
- 3:00–3:30 Random Positive-Only Projections: PPMI-Enabled Incremental Semantic Space Construction
Behrang QasemiZadeh and Laura Kallmeyer
- 3:30–4:00 **Break**
- Distributional semantics**
- 4:00–4:30 A Compositional-Distributional Semantic Model for Searching Complex Entity Categories
Juliano Efson Sales, Andre Freitas, Brian Davis, and Siegfried Handschuh
- 4:30–5:00 Approximating Givenness in Content Assessment through Distributional Semantics
Ramon Ziai, Kordula De Kuthy, and Detmar Meurers
- 5:00–5:30 Learning Embeddings to lexicalise RDF Properties
Laura Perez-Beltrachini and Claire Gardent
- 5:30–5:40 **Closing**

Message from the Program Committee Co-Chairs

The 2016 Conference on Computational Natural Language Learning is the twentieth in the series of annual meetings organized by SIGNLL, the ACL special interest group on natural language learning. CoNLL 2016 will be held on August 11-12, 2016, and is co-located with the 54th annual meeting of the Association for Computational Linguistics (ACL) in Berlin, Germany.

In order to accommodate papers with extended proofs and experimental material, CoNLL 2016 accepted only long papers, allowing 9 pages of content plus unlimited pages of references and supplementary material. We received 186 submissions in total, out of which 13 had to be rejected for formal reasons, and 21 were withdrawn by the authors. Of the remaining 149 papers, 30 papers were chosen to appear in the conference program, resulting in an overall acceptance rate of 20%. All accepted papers appear here in the proceedings.

As in previous years, CoNLL 2016 features a shared task, this year on Shallow Discourse Parsing. Papers accepted for the shared task are collected in a companion volume of CoNLL 2016.

To fit the paper presentations in a 2-day program, 21 long papers were selected for oral presentation, and 9 papers were presented as posters. The papers selected for oral presentation are distributed in six main sessions. Poster presenters were given the chance to present their poster in a short oral spotlight presentation.

For the first time, CoNLL 2016 announced a special topic on statistical natural language learning beyond linear models and convex optimization. The special topic was embraced by several authors and is reflected by the invited talks given by Jürgen Schmidhuber and Fernanda Ferreira.

We would like to thank all the authors who submitted their work to CoNLL 2016, as well as the program committee for helping us select the best papers out of many high-quality submissions. We are also grateful to our invited speakers, who graciously agreed to give talks at CoNLL.

Special thanks are due to the SIGNLL board members, Xavier Carreras and Julia Hockenmaier, for their valuable advice and assistance in putting together this year's program, and to Ben Verhoeven, for maintaining the CoNLL 2016 web page. We are grateful to the ACL organization for helping us with the program, proceedings and logistics. Finally, our gratitude goes to our sponsor, Google Inc., for supporting the best paper award at CoNLL 2016.

We hope you enjoy the conference!

Yoav Goldberg and Stefan Riezler
CoNLL 2016 conference co-chairs

CoNLL 2016: The SIGNLL Conference on Computational Natural Language Learning

Organizers: **Yoav Goldberg, Stefan Riezler**

Venue: 2016Loc

Thursday, August 11, 2016

9:00–9:10 **Opening**

9:10–10:30 **Session 1: Word-Level Semantics**

- Generating Sentences from a Continuous Space
Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio
- Identifying Temporal Orientation of Word Senses
Mohammed Hasanuzzaman, Gaël Dias, Stéphane Ferrari, Yann Mathet, and Andy Way
- Semi-supervised Clustering for Short Text via Deep Representation Learning
Zhiguo Wang, Haitao Mi, and Abraham Ittycheriah
- Neighborhood Mixture Model for Knowledge Base Completion
Dat Quoc Nguyen, Kairit Sirts, LiZhen Qu, and Mark Johnson

10:30–11:00 **Break**

11:00–12:20 **Session 2: Sentence-Level Semantics**

- context2vec: Learning Generic Context Embedding with Bidirectional LSTM
Oren Melamud, Jacob Goldberger, and Ido Dagan
- Learning to Jointly Predict Ellipsis and Comparison Structures
Omid Bakhtschinah, Alexis Cornelia Wellwood, and James Allen
- Event Embeddings for Semantic Script Modeling
Ashutosh Modi
- Beyond Centrality and Structural Features: Learning Information Importance for Text Summarization
Markus Zopf, Eneldo Loza Mencía, and Johannes Fürnkranz

12:20–12:50 **SIGNLL Business Meeting**

12:50–2:00 **Lunch Break**

2:00–3:40 **Session 3: Human Language Processing**

- Incremental Prediction of Sentence-final Verbs: Humans versus Machines
Alvin Grissom II, Naho Orita, and Jordan Boyd-Graber
- A Data-driven Investigation of Corrective Feedback on Subject Omission Errors in First Language Acquisition
Sarah Hiller and Raquel Fernandez
- Redefining part-of-speech classes with distributional semantic models
Andrey Kutuzov, Erik Veldal, and Lilja Øvrelid
- Analyzing Learner Understanding of Novel L2 Vocabulary
Rebecca Knowles, Adithya Renduchintala, Philipp Koehn, and Jason Eisner
- Modeling language evolution with codes that utilize context and phonetic features
Javad Nouri and Roman Yangarber

3:40–4:00 **Break**

4:00–5:15 **Keynote 1: Human Processing of Disfluent Speech: Basic Findings, Theoretical Approaches, and Implications for Natural Language Processing (Fernanda Ferreira)**

5:15–6:15 **Session 4: Sarcasm / Sentiment**

- Harnessing Sequence Labeling for Sarcasm Detection in Dialogue from TV Series ‘Friends’
Aditya Joshi, Vaibhav Tripathi, Pushpak Bhattacharyya, and Mark J Carman
- Leveraging Cognitive Features for Sentiment Analysis
Abhijit Mishra, Diptesh Kanodia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya
- Modelling Context with User Embeddings for Sarcasm Detection in Social Media
Silvio Amir, Byron C. Wallace, Hao Lyu, Paula Carvalho, and Mario J. Silva

Friday, August 12, 2016

9:00–9:15 **SIGNLL Steering Committee Address**

9:15–10:30 **Keynote 2: RNNaissance (Jürgen Schmidhuber)**

10:30–11:00 **Break**

11:00–12:40 **Session 5: Syntax, Named Entities, and Machine Translation**

- Learning when to trust distant supervision: An application to low-resource POS tagging using cross-lingual projection
Meng Fang and Trevor Cohn
- Greedy, Joint Syntactic-Semantic Parsing with Stack LSTMs
Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer, and Noah A. Smith
- Beyond Prefix-Based Interactive Translation Prediction
Jesús González-Rubio, Daniel Ortiz Martínez, Francisco Casacuberta, and Jose Miguel Benedí Ruiz
- Exploring Prediction Uncertainty in Machine Translation Quality Estimation
Daniel Beck, Lucia Specia, and Trevor Cohn
- Cross-Lingual Named Entity Recognition via Wikification
Chen-Tse Tsai, Stephen Mayhew, and Dan Roth

12:40–2:00 **Lunch Break**

2:00–3:30 **Session 6: Shared Task on Shallow Discourse Parsing**

3:30–4:00 **Break**

4:00–4:20 **Session 7: Spotlight Presentations**

4:20–5:20 **Session 8: Posters**

- Coreference in Wikipedia: Main Concept Resolution
Abbas Ghaddar and Phillippe Langlais
- Event Linking with Sentential Features from Convolutional Neural Networks
Sebastian Krause, Feiyu Xu, Hans Uszkoreit, and Dirk Weissenborn
- Joint Learning of the Embedding of Words and Entities for Named Entity Disambiguation
Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, and Yoshiyasu Takefuji
- Entity Disambiguation by Knowledge and Text Jointly Embedding
Wei Fang, Jianwen Zhang, Dilin Wang, Zheng Chen, and Ming Li
- Substring-based unsupervised transliteration with phonetic and contextual knowledge
Anoop Kunchukuttan, Pushpak Bhattacharyya, and Mitesh M. Khapra
- Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond
Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Caglar Gulcehre, and Bing Xiang
- Compression of Neural Machine Translation Models via Pruning
Abigail See, Minh-Thang Luong, and Christopher D. Manning
- Modelling the Usage of Discourse Connectives as Rational Speech Acts
Frances Yung, Kevin Duh, Taku Komura, and Yuji Matsumoto
- Semi-supervised Convolutional Networks for Translation Adaptation with Tiny Amount of In-domain Data
Boxing Chen and Fei Huang

5:20–6:00 **Session 9: Best Paper Awards + Closing**

8

Workshops

Thursday–Friday		
Governor's Square 17	First Conference on Machine Translation (WMT)	p.??
<u>Thursday</u>		
Governor's Square 12	1st Workshop on Representation Learning for NLP (RepL4NLP)	p.??
Director's Row E	SIGANN's Linguistic Annotation Workshop (LAW X 2016)	p.??
Director's Row I	12th Workshop on Multiword Expressions (MWE 2016)	p.??
Tower Court A	7th Workshop on Cognitive Aspects of Computational Language Learning	p.??
Director's Row H	14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology	p.??
Director's Row J	10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2016)	p.??
<u>Friday</u>		
Governors Square 11	5th Workshop on Vision and Language (VL'16)	p.??
Governors Square 17	15th Workshop on Biomedical Natural Language Processing	p.??
Director's Row E	SIGFSM Workshop on Statistical NLP and Weighted Automata	p.??
Director's Row I	1st Workshop on Evaluating Vector-Space Representations for NLP	p.??
Director's Row J	10th Web as Corpus Workshop	p.??
Governor's Square 11	6th NEWS Named Entities Workshop	p.??
Director's Row H	3rd Workshop on Argument Mining	p.??

Workshop 1: 1st Conference on Machine Translation

Organizers: *Barry Haddow, Philipp Koehn, Ondřej Bojar, Matthias Huck, Christian Federmann, Christof Monz, Matt Post, and Lucia Specia*

Venue: Governor's Square 17

Thursday, August 11, 2016

8:45–9:00 **Opening Remarks**

Session 1: Shared Tasks Overview Presentations I

9:00–9:20 **Shared Task: News Translation**

- Findings of the 2016 Conference on Machine Translation
Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Aurelie Neveol, Mariana Neves, Martin Popel, Matt Post, Raphael Rubino, Carolina Scarton, Lucia Specia, Marco Turchi, Karin Verspoor, and Marcos Zampieri

9:20–9:35 **Shared Task: IT-Domain Translation**

9:35–9:50 **Shared Task: Biomedical Translation**

9:50–10:10 **Shared Task: Metrics**

- Results of the WMT16 Metrics Shared Task
Ondřej Bojar, Yvette Graham, Amir Kamran, and Miloš Stanojević

10:10–10:30 **Shared Task: Tuning**

- Results of the WMT16 Tuning Shared Task
Bushra Jawaid, Amir Kamran, Miloš Stanojević, and Ondřej Bojar

10:30–11:00 **Coffee Break**

Session 2: Shared Tasks Poster Session I

11:00–12:30 **Shared Task: News Translation**

- LIMSIWMT'16: Machine Translation of News
Alexandre Allauzen, Lauriane Aufrant, Franck Burlot, Ophélie Lacroix, Elena Knyazeva, Thomas Lavergne, Guillaume Wisniewski, and François Yvon
- TÜBİTAK SMT System Submission for WMT2016
Emre Bektaş, Ertugrul Yilmaz, Coskun Mermer, and İlknur Durgar El-Kahlout
- ParFDA for Instance Selection for Statistical Machine Translation
Ergun Bicici
- Sheffield Systems for the English–Romanian WMT Translation Task
Frédéric Blain, Xingyi Song, and Lucia Specia
- MetaMind Neural Machine Translation System for WMT 2016
James Bradbury and Richard Socher
- NYU-MILA Neural Machine Translation Systems for WMT'16
Junyoung Chung, Kyunghyun Cho, and Yoshua Bengio
- The JHU Machine Translation Systems for WMT 2016
Shuoyang Ding, Kevin Duh, Huda Khayrallah, Philipp Koehn, and Matt Post
- Yandex School of Data Analysis approach to English–Turkish translation at WMT16 News Translation Task
Anton Dvorkovich, Sergey Gubanov, and Irina Galinskaya
- Hybrid Morphological Segmentation for Phrase-Based Machine Translation
Stig-Arne Grönroos, Sami Virpioja, and Mikko Kurimo
- The AFRL-MITLL WMT16 News-Translation Task Systems
Jeremy Gwinnup, Tim Anderson, Grant Erdmann, Katherine Young, Michael Kazi, Elizabeth Salesky, and Brian Thompson
- The Karlsruhe Institute of Technology Systems for the News Translation Task in WMT 2016
Thanh-Le Ha, Eunah Cho, Jan Niehues, Mohammed Mediani, Matthias Sperber, Alexandre Allauzen, and Alexander Waibel

- The Edinburgh/LMU Hierarchical Machine Translation System for WMT 2016
Matthias Huck, Alexander Fraser, and Barry Haddow
- The AMU-UEDIN Submission to the WMT16 News Translation Task: Attention-based NMT Models as Feature Functions in Phrase-based SMT
Marcin Junczys-Dowmunt, Tomasz Dwojak, and Rico Sennrich
- NRC Russian-English Machine Translation System for WMT 2016
Chi-ku Lo, Colin Cherry, George Foster, Darlene Stewart, Rabib Islam, Anna Kazantseva, and Roland Kuhn
- Merged bilingual trees based on Universal Dependencies in Machine Translation
David Mareček
- PROMT Translation Systems for WMT 2016 Translation Tasks
Alexander Molchanov and Fedor Bykov
- The QT21/HimL Combined Machine Translation System
Jan-Thorsten Peter, Tamer Alkhouri, Hermann Ney, Matthias Huck, Fabienne Braune, Alexander Fraser, Aleš Tamchyna, Ondřej Bojar, Barry Haddow, Rico Sennrich, Frédéric Blain, Lucia Specia, Jan Niehues, Alex Waibel, Alexandre Allauzen, Lauriane Aufrant, Franck Burlot, Thomas Lavergne, François Yvon, Joachim Daiber, and Mrcis Pinnis
- The RWTH Aachen University English-Romanian Machine Translation System for WMT 2016
Jan-Thorsten Peter, Tamer Alkhouri, Andreas Gutta, and Hermann Ney
- Abu-MaTran at WMT 2016 Translation Task: Deep Learning, Morphological Segmentation and Tuning on Character Sequences
Víctor M. Sánchez-Cartagena and Antonio Toral
- Edinburgh Neural Machine Translation Systems for WMT 16
Rico Sennrich, Barry Haddow, and Alexandra Birch
- The Edit Distance Transducer in Action: The University of Cambridge English-German System at WMT16
Felix Stahlberg, Eva Hasler, and Bill Byrne
- CUNI-LMU Submissions in WMT2016: Chimera Constrained and Beaten Aleš Tamchyna, Roman Sudarík, Ondřej Bojar, and Alexander Fraser
- Phrase-Based SMT for Finnish with More Data, Better Models and Alternative Alignment and Translation Tools
Jörg Tiedemann, Fabienne Cap, Jenna Kanerva, Filip Ginter, Sara Stygne, Robert Östling, and Marion Weller-Di Marco
- Edinburgh's Statistical Machine Translation Systems for WMT16
Philip Williams, Rico Sennrich, Maria Nadejde, Matthias Huck, Barry Haddow, and Ondřej Bojar
- PJAIT Systems for the WMT 2016
Krzysztof Wolk and Krzysztof Marasek

11:00–12:30 Shared Task: IT-domain Translation

- DFKI's system for WMT16 IT-domain task, including analysis of systematic errors
Eleftherios Avramidis
- ILLC-UvA Adaptation System (Scorpio) at WMT'16 IT-DOMAIN Task
Hoang Cuong, Stella Frank, and Khalil Sima'an
- Data Selection for IT Texts using Paragraph Vector
Mirela-Stefania Duma and Wolfgang Menzel
- SMT and Hybrid systems of the QTLeap project in the WMT16 IT-task
Rosa Gaudio, Gorka Labaka, Eneko Agirre, Petya Osenova, Kiril Simov, Martin Popel, Dieke Oele, Gertjan van Noord, Luís Gomes, João António Rodrigues, Steven Neale, João Silva, Andreia Querido, Nuno Rendeiro, and António Branco
- JU-USAAR: A Domain Adaptive MT System
Koushik Pahari, Alapan Kuila, Santanu Pal, Sudip Kumar Naskar, Sivaji Bandyopadhyay, and Josef van Genabith
- Dictionary-based Domain Adaptation of MT Systems without Retraining
Rudolf Rosa, Roman Sudarík, Michal Novák, Martin Popel, and Ondřej Bojar

11:00–12:30 Shared Task: Biomedical Translation

- English-Portuguese Biomedical Translation Task Using a Genuine Phrase-Based Statistical Machine Translation Approach
José Aires, Gabriel Lopes, and Luís Gomes
- The TALP—UPC Spanish—English WMT Biomedical Task: Bilingual Embeddings and Char-based Neural Language Model Rescoring in a Phrase-based System
Maria R. Costa-jussà, Cristina España-Bonet, Pranava Madhyastha, Carlos Escolano, and José A. R. Fonollosa

- LIMSI's Contribution to the WMT'16 Biomedical Translation Task
Julia Ive, Aurélien Max, and François Yvon
- IXA Biomedical Translation System at WMT16 Biomedical Translation Task
Olatz Perez-de-Viñaspre and Gorka Labaka

11:00–12:30 Shared Task: Metrics

- CobaltF: A Fluent Metric for MT Evaluation
Marina Fomicheva, Núria Bel, Lucia Specia, Iria da Cunha, and Anton Malinovskiy
- DTED: Evaluation of Machine Translation Structure Using Dependency Parsing and Tree Edit Distance
Martin McCaffery and Mark-Jan Nederhof
- chrF deconstructed: beta parameters and n-gram weights
Maja Popović
- CharacTer: Translation Edit Rate on Character Level
Weiyue Wang, Jan-Thorsten Peter, Hendrik Rosendahl, and Hermann Ney
- Extract Domain-specific Paraphrase from Monolingual Corpus for Automatic Evaluation of Machine Translation
Lilin Zhang, Zhen Weng, Wenyang Xiao, Jianyi Wan, Zhiming Chen, Yiming Tan, Maoxi Li, and Mingwen Wang

11:00–12:30 Shared Task: Tuning

- Particle Swarm Optimization Submission for WMT16 Tuning Task
Viktor Kocur and Ondřej Bojar

12:30–2:00 Lunch

Session 3: Invited Talk

2:00–3:30 Spence Green (Lilt): Interactive Machine Translation: From Research to Practice

3:30–4:00 Coffee Break

Session 4: Research Papers on Linguistic Modelling

- 4:00–4:20 Cross-language Projection of Dependency Trees with Constrained Partial Parsing for Tree-to-Tree Machine Translation
Yu Shen, Chenhua Chu, Fabien Cromieres, and Sadao Kurohashi
- 4:20–4:40 Improving Pronoun Translation by Modeling Coreference Uncertainty
Ngoc Quang Luong and Andrei Popescu-Belis
- 4:40–5:00 Modeling verbal inflection for English to German SMT
Anita Ramm and Alexander Fraser
- 5:00–5:20 Modeling Selectional Preferences of Verbs and Nouns in String-to-Tree Machine Translation
Maria Nadejde, Alexandra Birch, and Philipp Koehn
- 5:20–5:40 Modeling Complement Types in Phrase-Based SMT
Marion Weller-Di Marco, Alexander Fraser, and Sabine Schulte im Walde

Friday, August 12, 2016

Session 5: Shared Tasks Overview Presentations II

9:00–9:20 Shared Task: Cross-Lingual Pronoun Prediction

- Findings of the 2016 WMT Shared Task on Cross-lingual Pronoun Prediction
Liane Guillou, Christian Hardmeier, Preslav Nakov, Sara Stymne, Jörg Tiedemann, Yannick Versley, Mauro Cettolo, Bonnie Webber, and Andrei Popescu-Belis

9:20–9:40 Shared Task: Multimodal Machine Translation and Cross-Lingual Image Description

- A Shared Task on Multimodal Machine Translation and Crosslingual Image Description
Lucia Specia, Stella Frank, Khalil Sima'an, and Desmond Elliott

9:40–9:55 Shared Task: Bilingual Document Alignment

- Findings of the WMT 2016 Bilingual Document Alignment Shared Task
Christian Buck and Philipp Koehn

9:55–10:10 **Shared Task: Automatic Post-Editing**

10:10–10:30 **Shared Task: Quality Estimation**

10:30–11:00 **Coffee Break**

Session 6: Shared Tasks Poster Session II

11:00–12:30 **Shared Task: Cross-Lingual Pronoun Prediction**

- Cross-lingual Pronoun Prediction with Linguistically Informed Features
Rachel Bawden
- The Kyoto University Cross-Lingual Pronoun Translation System
Raj Dabre, Yevgeniy Puzikov, Fabien Cromieres, and Sadao Kurohashi
- Pronoun Prediction with Latent Anaphora Resolution
Christian Hardmeier
- It-disambiguation and source-aware language models for cross-lingual pronoun prediction
Sharid Loáiciga, Liane Guillou, and Christian Hardmeier
- Pronoun Language Model and Grammatical Heuristics for Aiding Pronoun Prediction
Ngoc Quang Luong and Andrei Popescu-Belis
- Cross-Lingual Pronoun Prediction with Deep Recurrent Neural Networks
Juhani Luotolahti, Jenna Kanerva, and Filip Ginter
- Pronoun Prediction with Linguistic Features and Example Weighing
Michał Novák
- Feature Exploration for Cross-Lingual Pronoun Prediction
Sara Stymne
- A Linear Baseline Classifier for Cross-Lingual Pronoun Prediction
Jörg Tiedemann
- Cross-lingual Pronoun Prediction for English, French and German with Maximum Entropy Classification
Dominikus Wetzel

11:00–12:30 **Shared Task: Multimodal Machine Translation and Cross-Lingual Image Description**

- Does Multimodality Help Human and Machine for Translation and Image Captioning?
Ozan Caglayan, Wälid Aransa, Yaxing Wang, Marc Masana, Mercedes García-Martínez, Feithi Bougares, Loïc Barrault, and Joost van de Weijer
- DCU-UvA Multimodal MT System Report
Iacer Calixto, Desmond Elliott, and Stella Frank
- Attention-based Multimodal Neural Machine Translation
Po-Yao Huang, Frederick Liu, Sz-Rung Shiang, Jean Oh, and Chris Dyer
- CUNI System for WMT16 Automatic Post-Editing and Multimodal Translation Tasks
Jindřich Libovický, Jindřich Helcl, Marek Tlustý, Ondřej Bojar, and Pavel Pecina
- WMT 2016 Multimodal Translation System Description based on Bidirectional Recurrent Neural Networks with Double-Embeddings
Sergio Rodríguez Guasch and Marta R. Costa-jussà
- SHEF-Multimodal: Grounding Machine Translation on Images
Kashif Shah, Josiah Wang, and Lucia Specia

11:00–12:30 **Shared Task: Bilingual Document Alignment**

- DOCAL - Vicomtech's Participation in the WMT16 Shared Task on Bilingual Document Alignment
Andoni Azpeitia and Thierry Etchegeoyen
- Quick and Reliable Document Alignment via TF/IDF-weighted Cosine Distance
Christian Buck and Philipp Koehn
- YODA System for WMT16 Shared Task: Bilingual Document Alignment
Aswartha Abhilash Dara and Yu-Chang Lin
- Bitextor's participation in WMT'16: shared task on document alignment
Miquel Esplà-Gomis, Mikel Forcada, Sergio Ortiz Rojas, and Jorge Ferrández-Tordera
- Bilingual Document Alignment with Latent Semantic Indexing
Ulrich Germann
- First Steps Towards Coverage-Based Document Alignment
Luís Gomes and Gabriel Pereira Lopes
- BAD LUCWMT 2016: a Bilingual Document Alignment Platform Based on Lucene
Laurent Jakubina and Phillippe Langlais

- Using Term Position Similarity and Language Modeling for Bilingual Document Alignment
Thanh Le, Hoa Trong Vu, Jonathan Oberländer, and Ondřej Bojar
- WMT2016: A Hybrid Approach to Bilingual Document Alignment
Sainik Mahata, Dipankar Das, and Santanu Pal
- English-French Document Alignment Based on Keywords and Statistical Translation
Marek Medved', Miloš Jakubíček, and Vojtěch Kovář
- The ILSP/ARC submission to the WMT 2016 Bilingual Document Alignment Shared Task
Vassilis Papavassiliou, Prokopis Prokopidis, and Stelios Piperidis
- Word Clustering Approach to Bilingual Document Alignment (WMT 2016 Shared Task)
Vadim Shchukin, Dmitry Khrustich, and Irina Galinskaya

11:00–12:30 Shared Task: Automatic Post-Editing

- The FBK Participation in the WMT 2016 Automatic Post-editing Shared Task
Rajen Chatterjee, José G. C. de Souza, Matteo Negri, and Marco Turchi
- Log-linear Combinations of Monolingual and Bilingual Neural Machine Translation Models for Automatic Post-Editing
Marcin Junczys-Dowmunt and Roman Grundkiewicz
- USAAR: An Operation Sequential Model for Automatic Statistical Post-Editing
Santanu Pal, Marcos Zampieri, and Josef van Genabith

11:00–12:30 Shared Task: Quality Estimation

- Bilingual Embeddings and Word Alignments for Translation Quality Estimation
Amal Abdelsalam, Ondřej Bojar, and Samhaa El-Beltagy
- SHEF-MIME: Word-level Quality Estimation Using Imitation Learning
Daniel Beck, Andreas Vlachos, Gustavo Paetzold, and Lucia Specia
- Referential Translation Machines for Predicting Translation Performance
Ergun Bicici
- UAlacant word-level and phrase-level machine translation quality estimation systems at WMT 2016
Miquel Esplà-Gomis, Felipe Sánchez-Martínez, and Mikel Forcada
- Recurrent Neural Network based Translation Quality Estimation
Hyun Kim and Jong-Hyeok Lee
- YSDA Participation in the WMT'16 Quality Estimation Shared Task
Anna Kozlova, Mariya Shmatova, and Anton Frolov
- USFD's Phrase-level Quality Estimation Systems
Varvara Logacheva, Frédéric Blain, and Lucia Specia
- Unbabel's Participation in the WMT16 Word-Level Translation Quality Estimation Shared Task
André F. T. Martins, Ramón Astudillo, Chris Hokamp, and Fabio Kepler
- SimpleNets: Quality Estimation with Resource-Light Neural Networks
Gustavo Paetzold and Lucia Specia
- Translation Quality Estimation using Recurrent Neural Network
Raj Nath Patel and Sasikumar M
- The UU Submission to the Machine Translation Quality Estimation Task
Oscar Sagemo and Sara Stygne
- Word embeddings and discourse information for Quality Estimation
Carolina Scarton, Daniel Beck, Kashif Shah, Karin Sim Smith, and Lucia Specia
- SHEF-LIUM-NN: Sentence level Quality Estimation with Neural Network Features
Kashif Shah, Fethi Bougares, Loïc Barrau, and Lucia Specia
- UGENT-LT3 SCATE Submission for WMT16 Shared Task on Quality Estimation
Arda Tezcan, Véronique Hoste, and Lieve Macken

12:30–2:00 Lunch

Session 7: Research Papers on Neural Translation

- 2:00–2:20 Alignment-Based Neural Machine Translation
Tamer Alkhouri, Gabriel Bretschner, Jan-Thorsten Peter, Mohammed Hethnawi, Andreas Gutai, and Hermann Ney
- 2:20–2:40 Neural Network-based Word Alignment through Score Aggregation
Joël Legrand, Michael Auli, and Ronan Collobert
- 2:40–3:00 Using Factored Word Representation in Neural Network Language Models
Jan Niehues, Thanh-Le Ha, Eunah Cho, and Alex Waibel
- 3:00–3:20 Linguistic Input Features Improve Neural Machine Translation
Rico Sennrich and Barry Haddow

3:30–4:00 **Coffee Break**

Session 8: Research Papers on Translation Models

4:00–4:20 A Framework for Discriminative Rule Selection in Hierarchical Moses
Fabiennne Braune, Alexander Fraser, Hal Daumé III, and Aleš Tamchyna

4:00–4:40 Fast and highly parallelizable phrase table for statistical machine translation
Nikolay Bogoychev and Hieu Hoang

4:00–5:00 A Comparative Study on Vocabulary Reduction for Phrase Table Smoothing
Yunsu Kim, Andreas Guta, Joern Wuebker, and Hermann Ney

5:00–5:20 Examining the Relationship between Preordering and Word Order Freedom in Machine Translation
Joachim Daiber, Miloš Stanojević, Wilker Aziz, and Khalil Sima'an

Workshop 2: 1st Workshop on Representation Learning for NLP

Organizers: *Phil Blunsom, Cho Kyunghyun, Shay Cohen, Edward Grefenstette, Karl Moritz Hermann, Laura Rimell, Jason Weston, and Scott Wen-tau Yih*

Venue: Governor's Square 12

Thursday, August 11, 2016

9:30–9:40 **Welcome and Opening Remarks**

9:40–10:30 **Keynote:** Katrin Erk (University of Texas at Austin)

10:30–11:00 **Coffee Break**

11:00–11:50 **Keynote:** Animashree Anandkumar (University of California, Irvine)

11:50–12:10 **Best Papers**

12:10–1:30 **Lunch Break**

1:30–2:20 **Keynote:** Hal Daumé III (University of Maryland)

2:20–3:10 **Keynote:** Raia Hadsell (DeepMind)

3:10–3:30 **Poster Session**

- Explaining Predictions of Non-Linear Classifiers in NLP
Leila Arras, Franziska Horn, Grégoire Montavon, Klaus-Robert Müller, and Wojciech Samek
- Joint Learning of Sentence Embeddings for Relevance and Entailment
Petr Baudiš, Silvestr Stanko, and Jan Šedivý
- A Joint Model for Word Embedding and Word Morphology
Kris Cao and Marek Rei
- On the Compositionality and Semantic Interpretation of English Noun Compounds
Corina Dima
- Functional Distributional Semantics
Guy Emerson and Ann Copestake
- Assisting Discussion Forum Users using Deep Recurrent Neural Networks
Jacob Hagedt P Suorra and Olof Mogren
- Adjusting Word Embeddings with Semantic Intensity Orders
Joo-Kyung Kim, Marie-Catherine de Marneffe, and Eric Fosler-Lussier
- Towards Abstraction from Extraction: Multiple Timescale Gated Recurrent Unit for Summarization
Minsoo Kim, Dennis Singh Moirangthem, and Minho Lee
- An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation
Jey Han Lau and Timothy Baldwin
- Quantifying the Vanishing Gradient and Long Distance Dependency Problem in Recursive Neural Networks and Recursive LSTMs
Phong Le and Willem Zuidema
- LSTM-Based Mixture-of-Experts for Knowledge-Aware Dialogues
Phong Le, Marc Dymetman, and Jean-Michel Renders
- Mapping Unseen Words to Task-Trained Embedding Spaces
Pranava Swaroop Madhyastha, Mohit Bansal, Kevin Gimpel, and Karen Livescu
- Multilingual Modal Sense Classification using a Convolutional Neural Network
Ana Marasović and Anette Frank
- Towards cross-lingual distributed representations without parallel text trained with adversarial autoencoders
Antonio Valerio Miceli Barone
- Decomposing Bilexical Dependencies into Semantic and Syntactic Vectors
Jeff Mitchell
- Learning Semantic Relatedness in Community Question Answering Using Neural Models
Henry Nassif, Mitra Mohtarami, and James Glass

- Learning Text Similarity with Siamese Recurrent Networks
Paul Neculoiu, Maarten Versteegh, and Mihai Rotaru
- A Two-stage Approach for Extending Event Detection to New Types via Neural Networks
Thien Huu Nguyen, Lisheng Fu, Kyunghyun Cho, and Ralph Grishman
- Parameterized context windows in Random Indexing
Tobias Norlund, David Nilsson, and Magnus Sahlgren
- Making Sense of Word Embeddings
Maria Pelevina, Nikolay Arefiev, Chris Biemann, and Alexander Panchenko
- Pair Distance Distribution: A Model of Semantic Representation
Yonatan Ramni, Oded Maimon, and Evgeni Khmelnitsky
- Measuring Semantic Similarity of Words Using Concept Networks
Gábor Recski, Eszter Iklódi, Katalin Pajkossy, and Andras Kornai
- Using Embedding Masks for Word Categorization
Stefan Rusevi, Traian Rebedea, and Stefan Trausan-Matu
- Sparsifying Word Representations for Deep Unordered Sentence Modeling
Prasanna Sattigeri and Jayaraman J. Thiagarajan
- Why “Blow Out”? A Structural Analysis of the Movie Dialog Dataset
Richard Searle and Megan Bingham-Walker
- Learning Word Importance with the Neural Bag-of-Words Model
Imran Sheikh, Irina Illina, Dominique Fohr, and Georges Linares
- A Vector Model for Type-Theoretical Semantics
Konstantin Sokolov
- Towards Generalizable Sentence Embeddings
Eleni Triantafillou, Jamie Ryan Kiros, Raquel Urtasun, and Richard Zemel
- Domain Adaptation for Neural Networks by Parameter Augmentation
Yusuke Watanabe, Kazuma Hashimoto, and Yoshimasa Tsuruoka
- Neural Associative Memory for Dual-Sequence Modeling
Dirk Weissenborn

3:30–4:00 **Poster Session Continues and Coffee break**

4:00–5:20 **Panel Discussion**

5:20–5:30 **Closing Remarks**

Workshop 3: 10th Linguistic Annotation Workshop

Organizers: *Nancy Ide, Adam Meyers, Katrin Tomanek, and Annemarie Friedrich*

Venue: Director's Row E

Thursday, August 11, 2016

Session 1: Opening and Invited Talk

9:00–9:10 **Opening Remarks**

9:10–10:05 **Invited talk: Assessing the Consistency and Use of “Common Sense” and Dependency Annotations (Marie-Catherine de Marneffe)**

10:05–10:30 Generating Disambiguating Paraphrases for Structurally Ambiguous Sentences
Manjujan Duan, Ethan Hill, and Michael White

10:30–11:00 **Coffee Break**

Session 2: Dependency Annotation and Discourse

11:00–11:25 Dependency Annotation Choices: Assessing Theoretical and Practical Issues of Universal Dependencies
Kim Gerdes and Sylvain Kahane

11:25–11:50 Conversion from Paninian Karakas to Universal Dependencies for Hindi Dependency Treebank
Juhu Tandon, Himani Chaudhry, Riyaz Ahmad Bhat, and Dipti Sharma

11:50–12:15 Different Flavors of GUM: Evaluating Genre and Sentence Type Effects on Multilayer Corpus Annotation Quality
Amir Zeldes and Dan Simonson

12:15–12:40 Annotating the discourse and dialogue structure of SMS message conversations
Nianwen Xue, Qishen Su, and Sooyoung Jeong

12:40–2:00 **Lunch Break**

Session 3: Evaluation of Agreement (Special Theme)

2:00–2:25 Evaluating Inter-Annotator Agreement on Historical Spelling Normalization
Marcel Böllmann, Stefanie Dipper, and Florian Petran

2:25–2:50 Filling in the Blanks in Understanding Discourse Adverbials: Consistency, Conflict, and Context-Dependence in a Crowdsourced Elicitation Task
Hannah Rohde, Anna Dickinson, Nathan Schneider, Christopher N. L. Clark, Annie Louis, and Bonnie Webber

2:50–3:05 **Poster boasters**

3:05–4:00 **Session 4: Poster Presentations**

- Building a Cross-document Event-Event Relation Corpus
Yu Hong, Tongtao Zhang, Tim O'Gorman, Sharone Horowitz-Hendler, Heng Ji, and Martha Palmer
- Annotating the Little Prince with Chinese AMRs
Bin Li, Yuan Wen, Weiguang QU, Lijun Bu, and Nianwen Xue
- Converting SynTagRus Dependency Treebank into Penn Treebank Style
Alex Luu, Sophia A. Malamud, and Nianwen Xue
- A Discourse-Annotated Corpus of Conjoined VPs
Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi
- Annotating Spelling Errors in German Texts Produced by Primary School Children
Ronja Laermann-Quante, Lukas Knichel, Stefanie Dipper, and Carina Betken
- Supersense tagging with inter-annotator disagreement
Héctor Martínez Alonso, Anders Johannsen, and Barbara Plank
- Filling in the Blanks in Understanding Discourse Adverbials: Consistency, Conflict, and Context-Dependence in a Crowdsourced Elicitation Task
Hannah Rohde, Anna Dickinson, Nathan Schneider, Christopher N. L. Clark, Annie Louis, and Bonnie Webber
- Comparison of Annotating Methods for Named Entity Corpora
Kanako Komiya, Masaya Suzuki, Tomoya Iwakura, Minoru Sasaki, and Hiroyuki Shinnou

- Addressing Annotation Complexity: The Case of Annotating Ideological Perspective in Egyptian Social Media
Heba Elfandy and Mona Diab
- Evaluating Inter-Annotator Agreement on Historical Spelling Normalization
Marcel Bollmann, Stefanie Dipper, and Florian Petran
- A Corpus of Preposition Supersenses
Nathan Schneider, Jena D. Hwang, Vivek Srikanth, Meredith Green, Abhijit Suresh, Kathryn Conger, Tim O'Gorman, and Martha Palmer
- Focus Annotation of Task-based Data: Establishing the Quality of Crowd Annotation
Kordula De Kuthy, Ramon Ziai, and Detmar Meurers
- Part of Speech Annotation of a Turkish-German Code-Switching Corpus
Özlem Çetinoğlu and Çağrı Çöltekin
- Phrase Generalization: a Corpus Study in Multi-Document Abstracts and Original News Alignments
Ariani Di-Felippo and Ani Nenkova
- Generating Disambiguating Paraphrases for Structurally Ambiguous Sentences
Manjuwan Duan, Ethan Hill, and Michael White
- Applying Universal Dependency to the Arapaho Language
Irina Wagner, Andrew Cowell, and Jena D. Hwang
- Creating a Novel Geolocation Corpus from Historical Texts
Grant DeLozier, Ben Wing, Jason Baldridge, and Scott Nesbit

3:30–4:00 **Poster presentations and coffee**

4:00–4:30 **Invited tutorial: How Item-Response Models Can Help us Take the Headache out of Annotation Projects (Dirk Hovy)**

4:30–5:15 **Discussion**

5:15–5:30 **Closing remarks: 10 years of LAW**

Workshop 4: 12th Workshop on Multiword Expressions

Organizers: *Valia Kordoni, Markus Egg, Kostadin Cholakov, Preslav Nakov, and Stella Markantonatou*

Venue: Director's Row I

Thursday, August 11, 2016

8:50–9:00 **Opening remarks**

Oral Session 1

- 9:00–9:30 Learning Paraphrasing for Multiword Expressions
Seid Muhie Yimam, Héctor Martínez Alonso, Martin Riedl, and Chris Biemann
- 9:30–10:00 Exploring Long-Term Temporal Trends in the Use of Multiword Expressions
Tal Daniel and Mark Last
- 10:00–10:30 Lexical Variability and Compositionality: Investigating Idiomaticity with Distributional Semantic Models
Marco Silvio Giuseppe Senaldi, Gianluca E. Lebani, and Alessandro Lenci

10:30–11:00 **Coffee Break**

Oral Session 2

- 11:00–11:20 Filtering and Measuring the Intrinsic Quality of Human Compositionality Judgments
Carlos Ramisch, Silvio Cordeiro, and Aline Villavicencio
- 11:20–11:40 Graph-based Clustering of Synonym Senses for German Particle Verbs
Moritz Wittmann, Marion Weller-Di Marco, and Sabine Schulte im Walde
- 11:40–12:00 Accounting ngrams and multi-word terms can improve topic models
Michael Nökel and Natalia Loukachevitch

12:00–1:00 **Invited Talk**

1:00–2:00 **Lunch**

2:00–2:40 **Poster Booster Session (5 minutes per poster)**

- Top a Splitter: Using Distributional Semantics for Improving Compound Splitting
Patrick Ziering, Stefan Müller, and Lonneke van der Plas
- Using Word Embeddings for Improving Statistical Machine Translation of Phrasal Verbs
Kostadin Cholakov and Valia Kordoni
- Modeling the Non-Substitutability of Multiword Expressions with Distributional Semantics and a Log-Linear Model
Mehdad Farahmand and James Henderson
- Phrase Representations for Multiword Expressions
Joël Legrand and Ronan Collobert
- Representing Support Verbs in FrameNet
Miriam R L Petrucci and Michael Ellsworth
- Inherently Pronominal Verbs in Czech: Description and Conversion Based on Treebank Annotation
Zdenka Uresova, Eduard Bejček, and Jan Hajic
- Using collocational features to improve automated scoring of EFL texts
Yves Bestgen
- A study on the production of collocations by European Portuguese learners
Angela Costa, Luísa Coheur, and Teresa Lino

2:40–3:30 **Poster Session**

3:30–4:00 **Coffee Break**

Oral Session 3

- 4:00–4:30 Extraction and Recognition of Polish Multiword Expressions using Wikipedia and Finite-State Automata
Paweł Chrząszcz
- 4:30–4:50 Impact of MWE Resources on Multiword Recognition
Martin Riedl and Chris Biemann
- 4:50–5:10 A Word Embedding Approach to Identifying Verb-Noun Idiomatic Combinations
Waseem Gharbieh, Virendra Bhavsar, and Paul Cook

5:10–5:20 **Closing Remarks**

Workshop 5: 7th Workshop on Cognitive Aspects of Computational Language Learning

Organizers: *Anna Korhonen, Alessandro Lenci, Thierry Poibeau, and Aline Villavicencio*

Venue: Tower Court A

Thursday, August 11, 2016

9:00–9:05 **Welcome and Opening Session**

Session 1 - Language in Clinical Conditions

9:05–9:35 Automated Discourse Analysis of Narrations by Adolescents with Autistic Spectrum Disorder
Michaela Regneri and Diane King

9:35–10:30 **Invited Speaker I**

10:30–11:00 **Coffee Break**

11:00–11:30 **Poster Session**

- Detection of Alzheimer's disease based on automatic analysis of common objects descriptions
Laura Hernandez-Dominguez, Edgar Garcia-Cano, Sylvie Ratté, and Gerardo Sierra Martínez
- Conversing with the elderly in Latin America: a new cohort for multimodal, multilingual longitudinal studies on aging
Laura Hernandez-Dominguez, Sylvie Ratté, Boyd Davis, and Charlene Pope
- Leveraging Annotators' Gaze Behaviour for Coreference Resolution
Joe Cheri, Abhijit Mishra, and Pushpak Bhattacharyya
- From alignment of etymological data to phylogenetic inference via population genetics
Javad Nouri and Roman Yangarber
- An incremental model of syntactic bootstrapping
Christos Christodoulopoulos, Dan Roth, and Cynthia Fisher

Session 2: Child Directed Language

11:30–12:00 Longitudinal Studies of Variation Sets in Child-directed Speech
Mats Wirén, Kristina Nilsson Björkenstam, Gintaré Grigonytė, and Elisabet Eir Cortes

12:00–12:30 Learning Phone Embeddings for Word Segmentation of Child-Directed Speech
Jianqiang Ma, Çağrı Çöltekin, and Erhard Hinrichs

12:30–2:00 **Lunch Break**

2:00–3:00 **Invited Talk II**

Session 3: Learning Artificial Languages

3:00–3:30 Generalization in Artificial Language Learning: Modelling the Propensity to Generalize
Raquel Garrido Alhama and Willem Zuidema

3:30–4:00 **Coffee Break**

Session 4: Language Acquisition

4:00–4:30 Explicit Causal Connections between the Acquisition of Linguistic Tiers:
Evidence from Dynamical Systems Modeling
Daniel Spokoyny, Jeremy Irvin, and Fermín Moscoso del Prado Martín

4:30–5:00 Modelling the informativeness and timing of non-verbal cues in parent-child interaction
Kristina Nilsson Björkenstam, Mats Wirén, and Robert Östling

Panel and Closing Session

5:00–5:30 **Panel**

5:30–5:35 **Closing Session**

Workshop 6: 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology

Organizers: *Micha Elsner and Sandra Kuebler*

Venue: Director's Row H

Thursday, August 11, 2016

Phonetics

9:00–9:30 Mining linguistic tone patterns with symbolic representation
SHUO ZHANG

9:30–10:30 Invited Talk (Colin Wilson)

10:30–11:00 Coffee Break

11:00–11:30 The SIGMORPHON 2016 Shared Task—Morphological Reinforcement
Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden

11:30–12:30 Shared task poster session

- Morphological reinlection with convolutional neural networks
Robert Östling
- EHU at the SIGMORPHON 2016 Shared Task. A Simple Proposal:
Grapheme-to-Phoneme for Inflection
Iñaki Alegria and Izaaskun Etxeberria
- Morphological Reinfection via Discriminative String Transduction
Garett Nicolai, Bradley Hauer, Adam St Arnaud, and Grzegorz Kondrak
- Morphological reinlection with conditional random fields and unsupervised features
Ling Liu and Lingshuang Jack Mao
- Improving Sequence to Sequence Learning for Morphological Inflection Generation:
The BIU-MIT Systems for the SIGMORPHON 2016 Shared Task for Morphological Reinfection
Roe Aharoni, Yoav Goldberg, and Yonatan Belinkov
- Evaluating Sequence Alignment for Learning Inflectional Morphology
David King
- Using longest common subsequence and character models to predict word forms
Alexey Sorokin
- MED: The LMU System for the SIGMORPHON 2016 Shared Task on Morphological Reinfection
Katharina Kann and Hinrich Schütze
- The Columbia University - New York University Abu Dhabi SIGMORPHON 2016
Morphological Reinfection Shared Task Submission
Dima Taji, Ramy Eskander, Nizar Habash, and Owen Rambow

12:00–2:00 Lunch Break

2:00–3:00 Workshop poster session

- Letter Sequence Labeling for Compound Splitting
Jianqiang Ma, Verena Henrich, and Erhard Hinrichs
- Automatic Detection of Intra-Word Code-Switching
Dong Nguyen and Leonie Cornips
- Read my points: Effect of animation type when speech-reading from EMA data
Kristy James and Martijn Wieling
- Predicting the Direction of Derivation in English Conversion
Max Kisselew, Laura Rimell, Alexis Palmer, and Sebastian Padó
- Morphological Segmentation Can Improve Syllabification
Garett Nicolai, Lei Yao, and Grzegorz Kondrak
- Towards a Formal Representation of Components of German Compounds
Thierry Declerck and Piroska Lendvai

Phonology

- 3:00–3:30 Towards robust cross-linguistic comparisons of phonological networks
Philippa Shoemark, Sharon Goldwater, James Kirby, and Rik Sarkar

3:30–4:00 Coffee Break

Morphology

- 4:00–4:30 Morphotactics as Tier-Based Strictly Local Dependencies
Alëna Aksënova, Thomas Graf, and Sedigheh Moradi
- 4:30–5:00 A Multilinear Approach to the Unsupervised Learning of Morphology
Anthony Meyer and Markus Dickinson
- 5:00–5:30 Inferring Morphotactics from Interlinear Glossed Text: Combining Clustering and Precision Grammars
Olga Zamaræva

Workshop 7: 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities

Organizers: *Nils Reiter, Beatrice Alex, and Kalliopi A. Zervanou*

Venue: Director's Row J

Thursday, August 11, 2016

Session 1

- 9:00–9:30 Brave New World: Uncovering Topical Dynamics in the ACL Anthology Reference Corpus Using Term Life Cycle Information
Anne-Kathrin Schumann
- 9:30–10:00 Analysis of Policy Agendas: Lessons Learned from Automatic Topic Classification of Croatian Political Texts
Mladen Karan, Jan Šnajder, Daniela Sirinic, and Goran Glavaš
- 10:00–10:30 Searching Four-Millenia-Old Digitized Documents: A Text Retrieval System for Egyptologists
Estibaliz Iglesias-Franjo and Jesús Vilares

Session 2

- 11:00–11:30 Old Swedish Part-of-Speech Tagging between Variation and External Knowledge
Yvonne Adesam and Gerlof Bouma
- 11:30–12:00 Code-Switching Ubique Est - Language Identification and Part-of-Speech Tagging for Historical Mixed Text
Sarah Schulz and Mareike Keller
- 12:00–12:30 Dealing with word-internal modification and spelling variation in data-driven lemmatization
Fabian Barteld, Ingrid Schröder, and Heike Zinsmeister

1:30–2:00 SIGHUM Business Meeting

Session 3

- 2:00–2:30 You Shall Know People by the Company They Keep: Person Name Disambiguation for Social Network Construction
Mariona Coll Ardanuy, Maarten van den Bos, and Caroline Sporleder
- 2:30–3:00 Deriving Players & Themes in the Regesta Imperii using SVMs and Neural Networks
Juri Opitz and Anette Frank

3:00–4:00 Poster Session

- Semi-automated annotation of page-based documents within the Genre and Multimodality framework
Tuomo Häippala
- Nomen Omen. Enhancing the Latin Morphological Analyser Lemlat with an Onomasticon
Marco Budassi and Marco Passarotti
- How Do Cultural Differences Impact the Quality of Sarcasm Annotation?: A Case Study of Indian Annotators and American Text
Aditya Joshi, Pushpak Bhattacharyya, Mark Carman, Jaya Saraswati, and Rajita Shukla
- Combining Phonology and Morphology for the Normalization of Historical Texts
Izaskun Exkeberria, Iñaki Alegria, Larraitz Uria, and Mans Hulden
- Towards Building a Political Protest Database to Explain Changes in the Welfare State
Çağil Sönmez, Arzucan Özgür, and Erdem Yörük
- An Assessment of Experimental Protocols for Tracing Changes in Word Semantics Relative to Accuracy and Reliability
Johannes Hellrich and Udo Hahn
- Universal Morphology for Old Hungarian
Eszter Simon and Veronika Vincze
- Automatic Identification of Suicide Notes from Linguistic and Sentiment Features
Annika Marie Schoene and Nina Dethlefs
- Towards a text analysis system for political debates
Dieu-Thu Le, Ngoc Thang Vu, and Andre Blessing

- Whodunit... and to Whom? Subjects, Objects, and Actions in Research Articles on American Labor Unions
Vilja Hulden

Session 4

- 4:00–4:30 An NLP Pipeline for Coptic
Amir Zeldes and Caroline T. Schroeder
- 4:30–5:00 Automatic discovery of Latin syntactic changes
Micha Elsner and Emily Lane
- 5:00–5:30 Information-based Modeling of Diachronic Linguistic Change: from Typicality to Productivity
Stefania Degaetano-Ortlieb and Elke Teich

Workshop 8: 5th Workshop on Vision and Language

Organizers: Anya Belz, Katerina Pastra, Erkut Erdem, and Krystian Mikolajczyk

Venue: Governors Square 17

Friday, August 12, 2016

9:00–9:30 **Opening Remarks**

9:30–10:00 Automatic Annotation of Structured Facts in Images

Mohamed Elhoseiny, Scott Cohen, Walter Chang, Brian Price, and Ahmed Elgammal

10:00–10:30 Combining Lexical and Spatial Knowledge to Predict Spatial Relations between Objects in Images

Manuela Hürlimann and Johan Bos

10:30–11:00 **Coffee Break**

11:00–12:00 **Invited talk: Yejin Choi**

12:00–12:30 Focused Evaluation for Image Description with Binary Forced-Choice Tasks

Micah Hodosh and Julia Höckenmaier

12:30–2:00 **Lunch Break**

2:00–2:30 Leveraging Captions in the Wild to Improve Object Detection

Mert Kilickaya, Nazli Ikizler-Cinbis, Erkut Erdem, and Aykut Erdem

2:30–3:30 **Quick-fire presentations for posters (5mins each)**

3:30–4:00 **Coffee Break**

4:00–5:30 **Poster Session**

- Natural Language Descriptions of Human Activities Scenes: Corpus Generation and Analysis
Nouf Alharbi and Yoshihiko Gotoh

- Interactively Learning Visually Grounded Word Meanings from a Human Tutor
Yanchao Yu, Arash Eshghi, and Oliver Lemon

- Pragmatic Factors in Image Description: The Case of Negations
Emiel van Miltenburg, Roser Morante, and Desmond Elliott

- Building a Bagpipe with a Bag and a Pipe: Exploring Conceptual Combination in Vision
Sandro Pezzelle, Ravi Shekhar, and Raffaella Bernardi

- Exploring Different Preposition Sets, Models and Feature Sets in Automatic Generation of Spatial Image Descriptions
Anja Belz, Adrian Muscat, and Brandon Birmingham

- Multi30K: Multilingual English-German Image Descriptions
Desmond Elliott, Stella Frank, Khalil Sima'an, and Lucia Specia

- “Look, some Green Circles!”: Learning to Quantify from Images
Ionut Sorodoc, Angeliki Lazaridou, Gemma Boleda, Aurélie Herbelot, Sandro Pezzelle, and Raffaella Bernardi

- Text2voronoi: An Image-driven Approach to Differential Diagnosis
Alexander Mehler, Tolga Ustu, and Wahed Hemati

- Detecting Visually Relevant Sentences for Fine-Grained Classification
Olivia Winn, Madhavan Kavanur Kidambi, and Smaranda Muresan

Workshop 9: 15th Workshop on Biomedical Natural Language Processing

Organizers: *Kevin B. Cohen, Dina Demner Fushman, Sophia Ananiadou, and Jun-ichi Tsujii*

Venue: Director's Row E

Friday, August 12, 2016

8:30–8:40 **Opening remarks**

Session 1: Entity extraction and representation

8:40–9:00 A Machine Learning Approach to Clinical Terms Normalization

Jose Castano, María Laura Gambarte, Hee Joon Park, Maria del Pilar Avila Williams, David Perez, Fernando Campos, Daniel Luna, Sonia Benitez, Hernan Berinsky, and Sofía Zanetti

9:00–9:20 Improved Semantic Representation for Domain-Specific Entities

Mohammad Taher Pilehvar and Nigel Collier

9:20–9:40 Identification, characterization, and grounding of gradable terms in clinical text

Chaitanya Shivade, Marie-Catherine de Marneffe, Eric Fosler-Lussier, and Albert M. Lai

9:40–10:00 Graph-based Semi-supervised Gene Mention Tagging

Golnar Sheikhab, Elizabeth Starks, Aly Karsan, Anoop Sarkar, and Inanc Birol

10:00–10:30 **Invited Talk: “The BioNLP-ST challenges on information extraction and knowledge acquisition in biology” – Robert Bossy and Jin-Dong Kim**

10:30–11:00 **Coffee Break**

Session 2: Event and Relation Extraction

11:00–11:20 Feature Derivation for Exploitation of Distant Annotation via Pattern Induction against Dependency Parses

Dayne Freitag and John Niekrasz

11:40–12:00 Inferring Implicit Causal Relationships in Biomedical Literature

Halil Kilicoglu

12:00–12:20 SnapToGrid: From Statistical to Interpretable Models for Biomedical Information Extraction

Marco A. Valenzuela-Escárcega, Gus Hahn-Powell, Dane Bell, and Mihai Surdeanu

12:20–12:40 Character based String Kernels for Bio-Entity Relation Detection

Ritambhara Singh and Yanjun Qi

12:40–2:00 **Lunch break**

Session 3: Disambiguation, Classification, and more

2:00–2:20 Disambiguation of entities in MEDLINE abstracts by combining MeSH terms with knowledge

Amy Siu, Patrick Ernst, and Gerhard Weikum

2:20–2:40 Using Distributed Representations to Disambiguate Biomedical and Clinical Concepts

Stephan Tulkens, Simon Suster, and Walter Daelemans

2:40–3:00 Unsupervised Document Classification with Informed Topic Models

Timothy Miller, Dmitriy Dligach, and Guergana Savova

3:00–3:20 Vocabulary Development To Support Information Extraction of Substance Abuse from Psychiatry Notes

Sumithra Velupillai, Danielle L Mowery, Mike Conway, John Hurdle, and Brent Kious

3:20–3:40 Syntactic analyses and named entity recognition for PubMed and PubMed Central — up-to-the-minute

Kai Hakala, Suwisa Kaewphan, Tapio Salakoski, and Filip Ginter

3:40–4:00 **Coffee Break**

4:00–4:30 **Invited Talk: “BioASQ: A challenge on large-scale biomedical semantic indexing and question answering” – Anastasia Krithara**

4:30–5:30 **Poster Session**

- Improving Temporal Relation Extraction with Training Instance Augmentation
Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard, and Guergana Savova
- Using Centroids of Word Embeddings and Word Mover's Distance for Biomedical Document Retrieval in Question Answering
Georgios-Ioannis Brokos, Prodromos Malakasiotis, and Ion Androutsopoulos
- Measuring the State of the Art of Automated Pathway Curation Using Graph Algorithms - A Case Study of the mTOR Pathway
Michael Spranger, Suseendra Palaniappan, and Samik Gosh
- Construction of a Personal Experience Tweet Corpus for Health Surveillance
Keyuan Jiang, Ricardo Calix, and Matrika Gupta
- Modelling the Combination of Generic and Target Domain Embeddings in a Convolutional Neural Network for Sentence Classification
Nut Limnopatnam and Nigel Collier
- PubTermVariants: biomedical term variants and their use for PubMed search
Lana Yeganova, Won Kim, Sun Kim, Rezarta Islamaj Doğan, Wanli Liu, Donald C Comeau, Zhiyong Lu, and W John Wilbur
- This before That: Causal Precedence in the Biomedical Domain
Gus Hahn-Powell, Dana Bell, Marco A. Valenzuela-Escárcega, and Mihai Surdeanu
- Syntactic methods for negation detection in radiology reports in Spanish
Viviana Cotik, Vanesa Stricker, Jorge Vivaldi, and Horacio Rodriguez
- How to Train good Word Embeddings for Biomedical NLP
Billy Chiu, Gamal Crichton, Anna Korhonen, and Sampo Pyysalo
- An Information Foraging Approach to Determining the Number of Relevant Features
Brian Connolly, Benjamin Glass, and John Pestian
- Assessing the Feasibility of an Automated Suggestion System for Communicating Critical Findings from Chest Radiology Reports to Referring Physicians
Brian E. Chapman, Danielle L Mowery, Evan Narasimhan, Neel Patel, Wendy Chapman, and Marta Heilbrun
- Building a dictionary of lexical variants for phenotype descriptors
Simon Kocabek and Tudor Groza
- Applying deep learning on electronic health records in Swedish to predict healthcare-associated infections
Olof Jacobson and Hercules Dalianis
- Identifying First Episodes of Psychosis in Psychiatric Patient Records using Machine Learning
Genevieve Gorrell, Sherifat Oduola, Angus Roberts, Tom Craig, Craig Morgan, and Rob Stewart
- Relation extraction from clinical texts using domain invariant convolutional neural network
Sunil Sahu, Ashish Anand, Krishnadev Oruganty, and Mahanandeshwar Gattu

Workshop 10: SIGFSM Workshop on Statistical NLP and Weighted Automata

Organizers: *Bryan Jurish, Andreas Maletti, Uwe Springmann, and Kay-Michael Würzner*

Venue: Director's Row I

Friday, August 12, 2016

Welcome and keynote

9:00–9:30 **Opening (Organizers)**

9:30–10:30 **Probabilistic Models of Related Strings (Jason Eisner)**

10:30–11:00 **Morning break**

Weighted tree automata

11:00–11:30 Equivalences between Ranked and Unranked Weighted Tree Automata via Binarization
Toni Dietze

11:30–12:00 Adaptive Importance Sampling from Finite State Automata
Christoph Teichmann, Kasimir Wansing, and Alexander Koller

12:00–12:30 Transition-based dependency parsing as latent-variable constituent parsing
Mark-Jan Nederhof

12:30–2:00 **Lunch break**

Weighted finite-state transducers

2:00–2:30 Distributed representation and estimation of WFST-based n-gram models
Cyril Allauzen, Michael Riley, and Brian Roark

2:30–3:00 Learning Transducer Models for Morphological Analysis from Example Inflections
Markus Fürstberg and Mans Hulden

3:00–3:30 Data-Driven Spelling Correction using Weighted Finite-State Methods
Miikka Silfverberg, Pekka Kauppinen, and Krister Lindén

3:30–4:00 **Afternoon break**

Various weighted automata

4:00–4:30 EM-Training for Weighted Aligned Hypergraph Bimorphisms
Frank Drewes, Kilian Gebhardt, and Heiko Vogler

4:30–5:00 On the Correspondence between Compositional Matrix-Space Models of Language and Weighted Automata
Shima Asaadi and Sebastian Rudolph

5:00–5:30 Pynini: A Python library for weighted finite-state grammar compilation
Kyle Gorman

Workshop 11: 1st Workshop on Evaluating Vector-Space Representations for NLP

Organizers: *Omer Levy, Felix Hill, Roi Reichart, Kyunghyun Cho, Anna Korhonen, Yoav Goldberg, and Antoine Bordes*

Venue: Director's Row J

Friday, August 12, 2016

9:00–9:15 Opening Remarks

9:15–10:00 Analysis Track

- Intrinsic Evaluation of Word Vectors Fails to Predict Extrinsic Performance
Billy Chiú, Anna Korhonen, and Sampo Pyysalo
- A critique of word similarity as a method for evaluating distributional semantic models
Miroslav Batchkarov, Thomas Kober, Jeremy Refin, Julie Weeds, and David Weir
- Issues in evaluating semantic spaces using word analogies
Tal Linzen

10:00–10:20 Proposal Track 1

- Evaluating Word Embeddings Using a Representative Suite of Practical Tasks
Neha Nayak, Gabor Angeli, and Christopher D. Manning
- Story Cloze Evaluator: Vector Space Representation Evaluation by Predicting What Happens Next
Nasrin Mostafazadeh, Lucy Vanderwende, Wen-tau Yih, Pushmeet Kohli, and James Allen

10:20–10:45 Coffee Break

10:45–12:30 Poster Session

10:45–11:00 Lightning Talks

- Problems With Evaluation of Word Embeddings Using Word Similarity Tasks
Manaal Faruqui, Yulia Tsvetkov, Pushpendre Rastogi, and Chris Dyer
- Intrinsic Evaluations of Word Embeddings: What Can We Do Better?
Anna Gladkova and Aleksandr Drozd
- Find the word that does not belong: A Framework for an Intrinsic Evaluation of Word Vector Representations
José Camacho-Collados and Roberto Navigli
- Capturing Discriminative Attributes in a Distributional Space: Task Proposal
Alicia Krebs and Denis Paperno
- An Improved Crowdsourcing Based Evaluation Technique for Word Embedding Methods
Farhana Ferdousi Liza and Marek Grzes
- Evaluation of acoustic word embeddings
Sahar Ghannay, Yannick Estève, Nathalie Camelin, and Paul Deleglise
- Evaluating Embeddings using Syntax-based Classification Tasks as a Proxy for Parser Performance
Arne Köhn
- Evaluating vector space models using human semantic priming results
Allison Ettinger and Tal Linzen
- Evaluating embeddings on dictionary-based similarity
Judit Ács and András Kornai
- Evaluating multi-sense embeddings for semantic resolution monolingually and in word translation
Gábor Borbély, Márton Makrai, Dávid Márk Nemeskey, and András Kornai
- Subsumption Preservation as a Comparative Measure for Evaluating Sense-Directed Embeddings
Ali Seyed
- Evaluating Informal-Domain Word Representations With UrbanDictionary
Naomi Saphra
- Thematic fit evaluation: an aspect of selectional preferences
Asad Sayeed, Clayton Greenberg, and Vera Demberg

12:30–2:00 **Lunch Break**

2:00–3:30 **Proposal Track 2: Word Representations**

- Improving Reliability of Word Similarity Evaluation by Redesigning Annotation Task and Performance Measure
Oded Avraham and Yoav Goldberg
- Correlation-based Intrinsic Evaluation of Word Vector Representations
Yulia Tsvetkov, Manaal Faruqui, and Chris Dyer
- Evaluating word embeddings with fMRI and eye-tracking
Anders Søgaard
- Defining Words with Words: Beyond the Distributional Hypothesis
Iuliana-Elena Parasca, Andreas Lukas Rauter, Jack Roper, Aleksandar Rusinov, Guillaume Bouchard, Sebastian Riedel, and Pontus Stenetorp

3:30–4:00 **Coffee Break**

4:00–5:30 **Proposal Track 3: Contextualized Representations**

- A Proposal for Linguistic Similarity Datasets Based on Commonality Lists
Dmitrijs Milajevs and Sascha Griffiths
- Probing for semantic evidence of composition by means of simple classification tasks
Alyson Ettinger, Ahmed Elgohary, and Philip Resnik
- SLEDDED: A Proposed Dataset of Event Descriptions for Evaluating Phrase Representations
Laura Rimell and Eva Maria Vecchi
- Sentence Embedding Evaluation Using Pyramid Annotation
Tal Baumel, Raphael Cohen, and Michael Elhadad

5:30–6:15 **Open Discussion**

6:15–6:30 **Best Proposal Awards**

Workshop 12: 10th Web as Corpus Workshop

Organizers: *Paul C. Cook, Stefan Evert, Roland Schäfer, and Egon Stemle*

Venue: Governor's Square 11

9:30–9:40 **Welcome and Introduction**

WAC-X Morning Session

9:40–10:00 Automatic Classification by Topic Domain for Meta Data Generation, Web Corpus Evaluation, and Corpus Comparison
Roland Schäfer and Felix Bildhauer

10:00–10:30 Efficient construction of metadata-enhanced web corpora
Adrien Barbaresi

10:30–11:00 **Coffee Break**

WAC-X Noon Session

11:00–11:30 Topically-focused Blog Corpora for Multiple Languages
Andrew Salway, Dag Elgesem, Knut Hofland, Øystein Reigem, and Lubos Steskal

11:30–12:00 The Challenges and Joys of Analysing Ongoing Language Change in Web-based Corpora: a Case Study
Anne Krause

12:00–12:30 Using the Web and Social Media as Corpora for Monitoring the Spread of Neologisms. The case of 'rapefugee', 'rapegee', and 'rapugee'.
Quirin Würschinger, Mohammad Fazleh Elahi, Desislava Zheкова, and Hans-Jörg Schmid

12:30–1:30 **Lunch Break**

EmpiriST Session

1:30–1:50 EmpiriST 2015: A Shared Task on the Automatic Linguistic Annotation of Computer-Mediated Communication and Web Corpora
Michael Beßwenger, Sabine Bartsch, Stefan Evert, and Kay-Michael Würzner

1:50–2:10 SoMaJo: State-of-the-art tokenization for German web and social media texts
Thomas Proisl and Peter Uhrig

2:10–2:30 UoS-(retraindistributionalsurface): Improving POS Tagging for OOV Words in German CMC and Web Data
Jakob Prange, Andrea Horbach, and Stefan Thater

WAC-X and EmpiriST Teaser Talks

2:30–2:35 Babler - Data Collection from the Web to Support Speech Recognition and Keyword Search
Gideon Mendels, Erica Cooper, and Julia Hirschberg

2:35–2:40 A Global Analysis of Emoji Usage
Nikola Ljubešić and Darja Fišer

2:40–2:45 Genre classification for a corpus of academic webpages
Erika Dalan and Serge Sharoff

2:45–2:50 On Bias-free Crawling and Representative Web Corpora
Roland Schäfer

2:55–3:00 EmpiriST: AIPHES - Robust Tokenization and POS-Tagging for Different Genres
Steffen Renus, Gerold Hintz, Chris Biemann, Christian M. Meyer, Darina Benikova, Judith Eckle-Kohler, Margot Mieskes, and Thomas Arnold

3:00–3:05 bot.zen @ EmpiriST 2015 - A minimally-deep learning PoS-tagger (trained for German CMC and Web data)
Egon Stemle

3:05–3:10 LTL-UDE @ EmpiriST 2015: Tokenization and PoS Tagging of Social Media Text
Tobias Horsmann and Torsten Zesch

3:10–4:30 **WAC-X and EmpiriST Poster Session**

Discussion Session

4:30–5:30 **WAC-X and EmpiriST Closing Discussion**

5:30–6:30 **Panel Discussion: Corpora, open science, and copyright reforms**

Workshop 13: 6th NEWS Named Entities Workshop

Organizers: *Rafael E. Banchs, Min Zhang, Haizhou Li, and A. Kumaran*

Venue: Director's Row H

Research Papers I

9:10–9:40 Leveraging Entity Linking and Related Language Projection to Improve Name Transliteration

Ying Lin, Xiaoman Pan, Aliya Deri, Heng Ji, and Kevin Knight

9:40–10:10 Multi-source named entity typing for social media

Reuth Vexler and Einat Minkov

10:10–10:30 Evaluating and Combining Name Entity Recognition Systems

Ridong Jiang, Rafael E. Banchs, and Haizhou Li

10:30–11:00 **Coffee Break**

Research Papers II

11:00–11:30 German NER with a Multilingual Rule Based Information Extraction System: Analysis and Issues

Anna Druzhkina, Alexey Leontyev, and Maria Stepanova

11:30–12:00 Spanish NER with Word Representations and Conditional Random Fields

Jenny Linet Copara Zea, Jose Eduardo Ochoa Luna, Camilo Thorne, and Goran Glavaš

12:00–12:30 Constructing a Japanese Basic Named Entity Corpus of Various Genres

Tomoya Iwakura, Kanako Komiya, and Ryuichi Tachibana

12:30–2:30 **Lunch Break**

2:30–3:30 **Keynote Speech**

2:30–3:30 Linguistic Issues in the Machine Transliteration of Chinese, Japanese and Arabic Names
Jack Halpern

3:30–4:00 **Coffee Break**

Shared Task on Name Entity Transliteration

4:00–4:10 Whitepaper of NEWS 2016 Shared Task on Machine Transliteration
Xiangyu Duan, Min Zhang, Haizhou Li, Rafael Banchs, and A. Kumaran

4:00–4:10 Report of NEWS 2016 Machine Transliteration Shared Task

Xiangyu Duan, Rafael Banchs, Min Zhang, Haizhou Li, and A. Kumaran

4:10–4:30 Applying Neural Networks to English-Chinese Named Entity Transliteration
Yan Shao and Joakin Nivre

4:30–4:50 Target-Bidirectional Neural Models for Machine Transliteration
Andrew Finch, Lemao Liu, Xiaolin Wang, and Eiichiro Sumita

4:50–5:10 Regulating Orthography-Phonology Relationship for English to Thai Transliteration
Binh Minh Nguyen, Hoang Gia Ngo, and Nancy F. Chen

5:10–5:20 Moses-based official baseline for NEWS 2016
Marta R. Costa-jussà

5:20–5:30 **Closing Remarks**

Workshop 14: 3rd Workshop on Argument Mining

Organizers: *Chris Reed, Kevin Ashley, Nancy Cardie, Claire Green, Iryna Gurevych, Georgios Litman, Diane and Petasis, Noam Slonim, and Vern Walker*

Venue: Director's Row H

Friday, August 12, 2016

9:00–9:10 **Welcome**

Session I

- 9:10–9:30 "What Is Your Evidence?" A Study of Controversial Topics on Social Media
Aseel Addawood and Masooda Bashir
- 9:30–9:50 Summarizing Multi-Party Argumentative Conversations in Reader Comment on News
Emma Barker and Robert Gaizauskas
- 9:50–10:10 Argumentative texts and clause types
Maria Becker, Alexis Palmer, and Anette Frank
- 10:10–10:30 Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews
Pavithra Rajendran, Danushka Bollegala, and Simon Parsons

10:30–11:00 **Coffee break**

Session II

- 11:00–11:20 The CASS Technique for Evaluating the Performance of Argument Mining
Rory Duthie, John Lawrence, Katarzyna Budzynska, and Chris Reed
- 11:20–11:40 Extracting Case Law Sentences for Argumentation about the Meaning of Statutory Terms
Jaromír Savelka and Kevin D. Ashley
- 11:40–12:00 Scrutinizable Feature Sets for Stance Classification
Angrosh Mandyia, Advaith Siddharthan, and Adam Wyner
- 12:00–12:15 Argumentation: Content, Structure, and Relationship with Essay Quality
Beata Beigman Klebanov, Christian Stab, Jill Burstein, Yi Song, Binod Gyawali, and Iryna Gurevych
- 12:15–12:30 Neural Attention Model for Classification of Sentences that Support Promoting/Suppressing Relationship
Yuta Koreeda, Toshihiko Yanase, Kohsuke Yanai, Misa Sato, and Yoshiki Niwa

12:30–2:00 **Lunch**

Session III

- 2:00–2:20 Towards Feasible Guidelines for the Annotation of Argument Schemes
Elena Musi, Debanjan Ghosh, and Smaranda Muresan
- 2:20–2:40 Identifying Argument Components through TextRank
Georgios Petasis and Vangelis Karkaletsis
- 2:40–3:00 Rhetorical structure and argumentation structure in monologue text
Andreas Peldszus and Manfred Stede
- 3:00–3:15 Recognizing the Absence of Opposing Arguments in Persuasive Essays
Christian Stab and Iryna Gurevych
- 3:15–3:30 Expert Stance Graphs for Computational Argumentation
Orith Toledo-Ronen, Roy Bar-Haim, and Noam Slonim

3:30–4:00 **Coffee break**

Session IV - Debating Technologies and the Unshared Task Panel

- 4:00–4:20 Fill the Gap! Analyzing Implicit Premises between Claims from Online Debates
Filip Boltuzic and Jan Šnajder
- 4:20–4:40 Summarising the points made in online political debates
Charlie Egan, Advaith Siddharthan, and Adam Wyner
- 4:40–5:00 What to Do with an Airport? Mining Arguments in the German Online Participation Project Tempelhofer Feld
Matthias Liebeck, Katharina Esau, and Stefan Conrad

5:00–5:25 Panel Discussion: Unshared Task

- Unshared task: (Dis)agreement in online debates
Maria Skeppstedt, Magnus Sahlgren, Carita Paradis, and Andreas Kerren
- Unshared Task at the 3rd Workshop on Argument Mining: Perspective Based Local Agreement and Disagreement in Online Debate
Chantal van Son, Tommaso Caselli, Antske Fokkens, Isa Maks, Roser Morante, Lora Aroyo, and Piek Vossen
- A Preliminary Study of Disputation Behavior in Online Debating Forum
Zhongyu Wei, Yandi Xia, Chen Li, Yang Liu, Zachary Stallbohm, Yi Li, and Yang Jin

5:25–5:30 Closing Remarks

5:30–5:45 Close

Anti-harassment policy

The open exchange of ideas, the freedom of thought and expression, and respectful scientific debate are central to the aims and goals of a NAACL conference. These require a community and an environment that recognizes the inherent worth of every person and group, that fosters dignity, understanding, and mutual respect, and that embraces diversity. For these reasons, NAACL is dedicated to providing a harassment-free experience for participants at our events and in our programs.

Harassment and hostile behavior are unwelcome at any NAACL conference. This includes: speech or behavior (including in public presentations and on-line discourse) that intimidates, creates discomfort, or interferes with a person's participation or opportunity for participation in the conference. We aim for NAACL conferences to be an environment where harassment in any form does not happen, including but not limited to: harassment based on race, gender, religion, age, color, national origin, ancestry, disability, sexual orientation, or gender identity. Harassment includes degrading verbal comments, deliberate intimidation, stalking, harassing photography or recording, inappropriate physical contact, and unwelcome sexual attention.

It is the responsibility of the community as a whole to promote an inclusive and positive environment for our scholarly activities. In addition, any participant who experiences harassment or hostile behavior may contact any current member of the NAACL Board or contact Priscilla Rasmussen, who is usually available at the registration desk of the conference. Please be assured that if you approach us, your concerns will be kept in strict confidence, and we will consult with you on any actions taken.

The NAACL board members are listed at:

<http://naacl.org/officers/officers-2015.html>

The full policy and its implementation is defined at:

<http://naacl.org/policies/anti-harassment.html>

10

Local Guide

This guide was written by Kevin Cohen. For the most up-to-date version, please visit <http://bretonnel.com/2015/04/29/visiting-denver-for-naacl2015/>.

The North American Association for Computational Linguistics annual meeting (NAACL 2015) will be held in Denver, Colorado this year. Here are some things that I think might be useful or enjoyable for visiting computational linguists, natural language processing people, and the like. I'm not going to talk about the mountains, Red Rocks, or any of that kind of stuff—you can find that in tourist guides, hotel propaganda, and pretty much anywhere else. These are some of the things that make life in Denver bearable, and that I don't think you'll hear about elsewhere.

Airport The Denver International Airport looks quite distinctive. Opinions differ as to whether it is meant to look like teepees, the Rocky Mountains, or what. It features in a number of conspiracy theories, which mainly claim that it is built over an underground complex that will be the seat of the government of the New World Order. On the way from the airport to Denver, be sure to look for the Demon Horse statue. We call it the Demon Horse for two reasons: (1) it has glowing red eyes, and (2) during its construction, the head fell off and crushed the artist, killing him.

Bookstores Here's a link to a web page listing ten good or great independent Denver bookstores: www.westword.com/arts/the-ten-best-bookstores-in-denver-6659360

Bars The classic Denver bar that no one else knows about is El Chapultepec. Either arrive early, or be prepared to stand all evening. You can reach it from the conference hotel with a ride down the 16th Street Mall free shuttle and a short walk through a lively neighborhood. The LoDo area has many bars that are quite busy on weekend nights. Use caution around the time that the bars close. Again, you can reach LoDo quite easily from the conference hotel via the free 16th Street Mall shuttle.

Restaurants A rare hippie restaurant treat in Denver is the Mercury Cafe, known to us locally as "The Merc." It's a step back into the 1960s, sorta. Take a cab there, or the light rail—don't try to walk from the conference hotel, as the neighborhood is not always safe.

Marijuana Marijuana is legal here under state law. You can find it easily; the stores (usually known as "dispensaries") are typically marked with a green cross or a green marijuana leaf. However, it is NOT legal under federal law—if you are not an American citizen, don't take a chance with this. It's a legal gray zone, and people do occasionally get burnt. Also, like alcohol, it is not legal to consume it in public. (And, no: I don't indulge!)

Mexican food The Denver population is 30% Hispanic, and we have fantastic Mexican food here—also Salvadoran and some Peruvian, if you don't mind leaving the area of the conference hotel to find it. Mexican food is an integral part of American food in this part of the country. A good place to get it is Real de Minas,

on Colfax. Avoid the cheese-smothered burrito platters and have something that's actually Mexican, like tacos de carne asada, or ribs (costillas) in green chile sauce. You can get there on the number 15 bus—more on that below.

Decadent snacks Voodoo Doughnuts is an import from Portland, Oregon—some of you may remember it from ACL 2011. Truly amazing doughnuts—be prepared to stand in line. You can get there on the number 15 bus from the conference hotel.

Local beers Denver has a lot of microbreweries, and many good local beers. One of the main favorites is Fat Tire (which is now nationally distributed, so you may have had it before).

The Number 15 bus The number 15 bus goes up and down Colfax Avenue, allegedly the longest street in America, and probably one of the sleaziest. (Colfax runs quite close to the conference hotel.) Everyone in Denver has a story about the number 15 bus, typically involving a drunk, a drug addict, or vomit. It's actually pretty safe, although you should be careful on Colfax at night, as you would in any big city anywhere in the world. The last stop on the eastbound leg of the route (away from the mountains) is the Anschutz/Fitzsimons medical campus. Stop by the Biomedical Text Mining Group in the Center for Computational Bioscience for one of the best views of downtown Denver and the mountains that you'll find.

Safety Denver is a pretty safe city. Aurora is not quite as safe, particularly in the older parts of town. In general, you should be aware of your surroundings in the evening, as you would be in any big city.

Exterior of Denver Art Museum by Jeff Wells, courtesy Denver Art Museum. Other photos by Steve Trice.

Art at Every Angle



"The Mile High City is remaking itself as a world capital of art and architecture." **Sunset Magazine**

The spectacular **Denver Art Museum** (DAM) is the largest art museum between Kansas City and the West Coast. Designed by world-renowned architect Daniel Libeskind, the Hamilton Building is a work of art in itself. Since it was completed in 2006, the DAM has become a bona fide arts-world icon. "The building is smashing," raved *Newsweek*. The addition doubled the size of the museum and allows the DAM to host major touring exhibits, including "King Tut," through January 9, 2011.

Inside, the 356,000-square-foot DAM is filled with classics by Monet, Picasso and Matisse, alongside more modern works by Warhol and O'Keefe. All in all, the museum contains a collection numbering more than 68,000 works from around the world, including intriguing pieces from Africa and pre-Columbian America. The DAM is also home to one of the greatest collections of Western American art in the world.

Denver Performing Arts Complex (DPAC)



Tony Award®-winning theater, Broadway touring productions, contemporary dance and ballet, magnificent chorales, a major symphony orchestra, internationally acclaimed opera – it's all happening at this 12-acre, four-square-block complex in the heart of downtown Denver. **DPAC** is the second-largest arts complex in the nation, with 10 performance spaces seating 10,000 people, all connected by an 80-foot-tall glass ceiling.

What's GOING ON?

To find everything happening at Denver's cultural facilities and art galleries, visit: www.Denver365.com. Here you'll find the latest on theater, dance, symphony, opera, art show openings, film, museums, kid-friendly activities and annual cultural events, from the Cherry Creek Arts Festival to Denver Arts Week.

Art Districts



Hip galleries, world-class museums and First Friday artwalks – Denver's art districts are the pulse of the Mile High City's creative community. **Downtown Denver** has a variety of art galleries, many of which are found in LoDo with beautifully restored brick warehouses, and **Cherry Creek North** features galleries amidst its boutiques and restaurants. More Denver neighborhoods, including **Highlands**, **Navajo Street**, **Old South Pearl** and **Old South Gaylord Street**, are home to eclectic galleries and studios. The **ArtDistrict on Santa Fe** boasts more than 40 galleries, shops, and restaurants.

The **Golden Triangle Museum District** is home to eight museums, in addition to more than 50 galleries, fine art studios and specialty stores. Just north of downtown, the **River North Art District**, which goes by the catchy nickname **RiNo**, is where art is made. The District is also earning a name for itself amongst in-the-know creative types, thanks to a see-and-be-seen First Friday celebration, with galleries and shops staying open late.

Museum of Contemporary Art Denver (MCA)



Denver's cutting-edge arts world was given a 27,000-square-foot home in 2007 with the completion of the David Adjaye-designed **Museum of Contemporary Art Denver** in Lower Downtown (LoDo). This lovely, modern building – with five intimate gallery spaces and a panoramic rooftop café – houses a constantly refreshed set of exhibits. No visit here is ever the same.



Climate Chart

Month	Temperature Max/Min	% of Humidity AM/PM	Precipitation Inches	Sunshine Noon %
JANUARY	43.2/16.1 F	6.21/-8.8 C	63/49	.50
FEBRUARY	46.6/20.2 F	8.1/-6.6 C	67/44	.57
MARCH	52.2/25.8 F	11.2/-3.4 C	68/40	1.28
APRIL	61.8/34.5 F	16.6/-1.4 C	67/35	1.71
MAY	70.8/43.6 F	21.6/6.4 C	70/38	2.40
JUNE	81.4/52.4 F	21.6/6.4 C	65/35	1.79
JULY	88.2/58.6 F	31.2/14.8 C	68/34	1.91
AUGUST	85.5/56.9 F	29.7/13.8 C	69/35	1.51
SEPTEMBER	76.9/47.6 F	24.9/8.7 C	69/34	1.24
OCTOBER	66.3/36.4 F	19.1/2.4 C	65/35	.98
NOVEMBER	52.5/25.4 F	11.4/-3.7 C	68/48	.87
DECEMBER	44.5/17.4 F	6.9/-8.1 C	65/51	.64
ANNUAL AVERAGE	64.2/36.2 F	17.9/2.3 C	67/40	15.4
				70

Source: U.S. National Oceanic and Atmospheric Administration

Denver International Airport

Denver International Airport is the largest airport in the nation, covering 53 square miles – large enough to hold both Dallas-Fort Worth and Chicago O'Hare airports. The airport is a short 24-mile drive from downtown Denver, making it easy to access the Mile High City.



High Altitude Tips



Drink lots of water – even if you're not thirsty.

Use sunscreen: There is 25 percent less protection from the sun's rays at Denver's 5,280-foot elevation.

In mile-high Denver's rarefied air, a golf ball goes 10 percent farther. So does a cocktail. Drink alcohol in moderation.

Take some time to look at the sky. There is less water vapor in the air a mile above sea level, so the sky really is bluer.

The Napa Valley of Beer



Denver is the "Napa Valley of Beer," brewing more than 90 different beers daily. Coors Brewery is the largest single brewing site in the world and the Great American Beer Festival, held each September, is the nation's largest celebration of suds with 1,900 beers to sample. Denver Beer Fest overlaps from September 10-19. Go to DenverBeerFest.com for more information.



5 FREE things to do in Denver

- 1 Stand a "mile high" on the steps of the State Capitol – 5,280 feet above sea level.
- 2 Ride the FREE hybrid shuttle bus on the mile-long 16th Street Mall.
- 3 Tour the U.S. Mint, where more than 50 million coins are made each year.
- 4 Hike the spectacular trails at Red Rocks Amphitheatre & Park.
- 5 Drink a free beer (for those over 21) after touring the world's largest brewing site, Coors Brewery, in Golden.



Downtown - 16th Street Mall



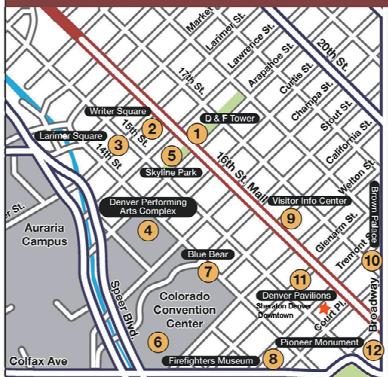
DOWNTOWN DENVER

The 16th Street Mall



The heart of downtown Denver is the 16th Street Mall, a mile-long pedestrian promenade lined with 200 trees, 50,000 flowers, outdoor cafés and inviting shops. Designed by famed architect I.M. Pei, the mall features a pattern in colored granite that resembles the design on a diamondback rattlesnake when viewed from above.

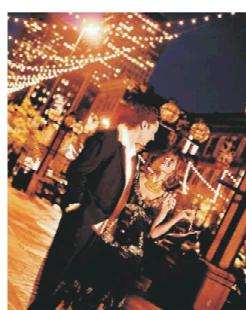
Free hybrid buses leave either end of the mall as often as every 90 seconds, stopping on every corner. The pedestrian portion of the 16th Street Mall continues over two footbridges to Commons Park on the South Platte River and to the Highlands neighborhood.



1 DANIELS & FISHER TOWER. Known by locals as the D&F Tower, this was the highest building west of the Mississippi River when it opened in 1909 and the third highest structure in the U.S. It is a $\frac{3}{4}$ -scale model of the celebrated bell tower of San Marco in Venice.



2 WRITER SQUARE. This pleasant plaza named after a local Denver family, is lined with flower baskets and outdoor sculptures, offering quiet cafés and shops.



3 LARIMER SQUARE. Downtown Denver's trendiest block is also one of its most historic. Beautiful Victorian brick-and-stone buildings now house chic eateries, one-of-a-kind galleries and clothing boutiques, nightspots, a comedy club, and some of Denver's top restaurants. This is the perfect place to watch a sunset from a sidewalk cafe.

Downtown Map: 1. D&F Tower by Riccardo Sartori; 2. Writer Square by Steve Cicaliello; 3. Blue Bear by Rich Grana; 4. Denver Pavilions by Steve Cicaliello; 5. Colorado Convention Center by Steve Cicaliello; 6. Firefighters Museum by Steve Cicaliello.



4 DENVER PERFORMING ARTS COMPLEX. The second-largest performing arts complex in the nation offers 10 venues seating more than 10,000 people for Tony® Award-winning theater, opera, symphony and ballet. The new \$92-million Ellie Caulkins Opera House opened in 2005 and Boettcher Concert Hall is undergoing a massive refurbishing. Ninety-minute backstage tours are available, visiting everything from the dressing rooms to costume and scenery shops.



8 FIREFIGHTERS MUSEUM. The hottest place in town tells the story of Denver's firefighters, from 1866 to present, featuring hands-on fun for children and adults with a unique gift shop.

9 DENVER VISITOR INFORMATION CENTER. Pop into our one-stop-shop for everything Denver – hundreds of free brochures and maps, personal assistance in making all your plans, a TicketMaster outlet and a gift shop.



10 BROWN PALACE HOTEL & SPA. Built in 1892, Denver's grand dame hotel features a dramatic eight-story atrium topped by a stained-glass ceiling. Enjoy afternoon tea in the lobby or take a historic walking tour of the hotel every Wednesday and Saturday at 3 p.m. Private tours can also be scheduled for other days or groups.



6 COLORADO CONVENTION CENTER. Ranked by meeting planners as one of the top 5 convention centers in the nation, this massive facility offers 592,000 square feet of exhibit space – a 14-acre room. Two massive lobbies feature more than 4.5 acres of glass and offer spectacular views, while the facility has 62 meeting rooms and a beautiful 5,000-seat theater.

5 SKYLINE PARK. This three-block-long oasis in the heart of the city offers a pleasant green space to relax. Concerts and special events take place here throughout the year.



7 BLUE BEAR. One of the most popular public art statues in Denver is Lawrence Argent's 40-foot high Blue Bear, "I See What You Mean," which appears to be peeking into the convention center. The 10,000-pound bear is made up of 4,000 interlocking triangles.



11 DENVER PAVILIONS. Check out two square blocks and three levels of shopping, dining and entertainment with Niketown, Barnes & Noble bookstore, Lucky Strike Lanes, Hard Rock Café and more.



12 PIONEER MONUMENT & FOUNTAIN. The bronze figure on the rearing horse pointing the way west is Christopher "Kit" Carson, the legendary explorer and scout who played an important role in developing the Rocky Mountain West.

Climate Chart

Month	Temperature Max/Min	% of Humidity AM/PM	Precipitation Inches	Sunshine Noon %
JANUARY	43.2/16.1 F	6.21/-8.8 C	63/49	.50
FEBRUARY	46.6/20.2 F	8.1/-6.6 C	67/44	.57
MARCH	52.2/25.8 F	11.2/-3.4 C	68/40	1.28
APRIL	61.8/34.5 F	16.6/-1.4 C	67/35	1.71
MAY	70.8/43.6 F	21.6/6.4 C	70/38	2.40
JUNE	81.4/52.4 F	21.6/6.4 C	65/35	1.79
JULY	88.2/58.6 F	31.2/14.8 C	68/34	1.91
AUGUST	85.5/56.9 F	29.7/13.8 C	69/35	1.51
SEPTEMBER	76.9/47.6 F	24.9/8.7 C	69/34	1.24
OCTOBER	66.3/36.4 F	19.1/2.4 C	65/35	.98
NOVEMBER	52.5/25.4 F	11.4/-3.7 C	68/48	.87
DECEMBER	44.5/17.4 F	6.9/-8.1 C	65/51	.64
ANNUAL AVERAGE	64.2/36.2 F	17.9/2.3 C	67/40	15.4
				70

Source: U.S. National Oceanic and Atmospheric Administration

Denver International Airport

Denver International Airport is the largest airport in the nation, covering 53 square miles – large enough to hold both Dallas-Fort Worth and Chicago O'Hare airports. The airport is a short 24-mile drive from downtown Denver, making it easy to access the Mile High City.



High Altitude Tips



Drink lots of water – even if you're not thirsty.

Use sunscreen: There is 25 percent less protection from the sun's rays at Denver's 5,280-foot elevation.

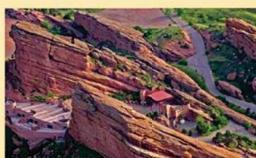
In mile-high Denver's rarefied air, a golf ball goes 10 percent farther. So does a cocktail. Drink alcohol in moderation.

Take some time to look at the sky. There is less water vapor in the air a mile above sea level, so the sky really is bluer.

The Napa Valley of Beer



Denver is the "Napa Valley of Beer," brewing more than 90 different beers daily. Coors Brewery is the largest single brewing site in the world and the Great American Beer Festival, held each September, is the nation's largest celebration of suds with 1,900 beers to sample. Denver Beer Fest overlaps from September 10-19. Go to DenverBeerFest.com for more information.



5 FREE things to do in Denver

1 Stand a "mile high" on the steps of the State Capitol – 5,280 feet above sea level.

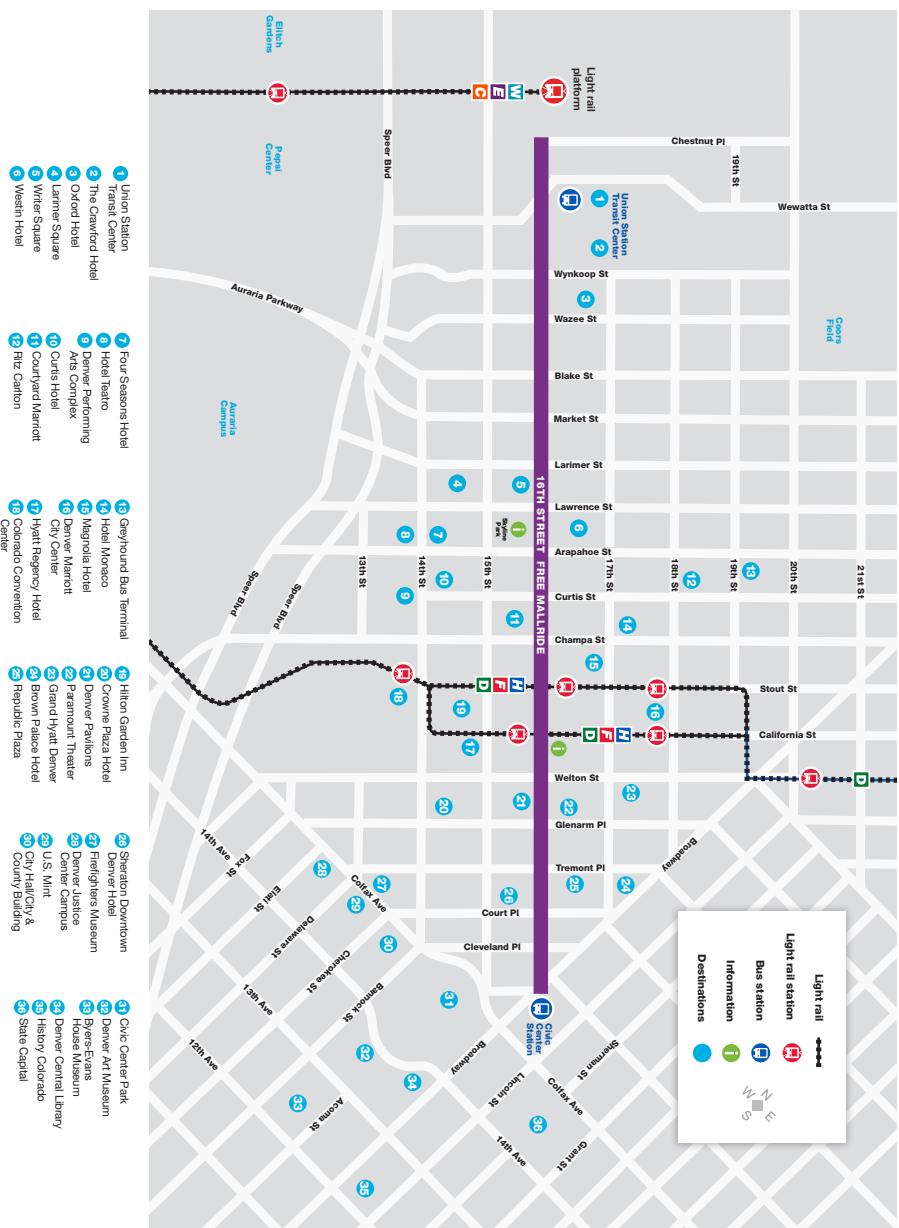


at Red Rocks Amphitheatre & Park.

2 Ride the FREE hybrid shuttle bus on the mile-long 16th Street Mall.

4 Hike the spectacular trails

5 Drink a free beer (for those over 21) after touring the world's largest brewing site, Coors Brewery, in Golden.



NAACL gratefully acknowledges the following sponsors for their support:

Platinum

Bloomberg

Gold



**Goldman
Sachs**



Silver



Bronze

YAHOO! LABS

Supporter



Student Research Workshop



Wireless



MACHINEZONE