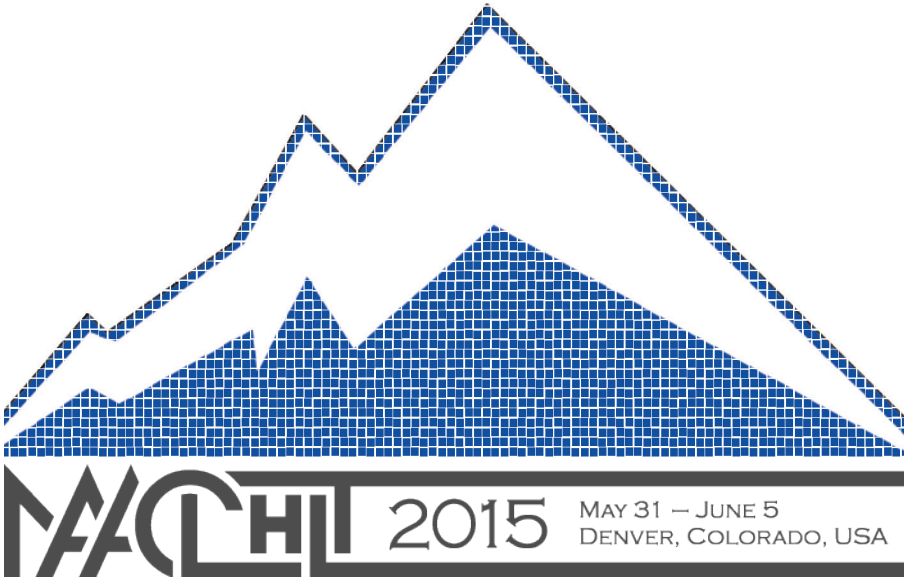
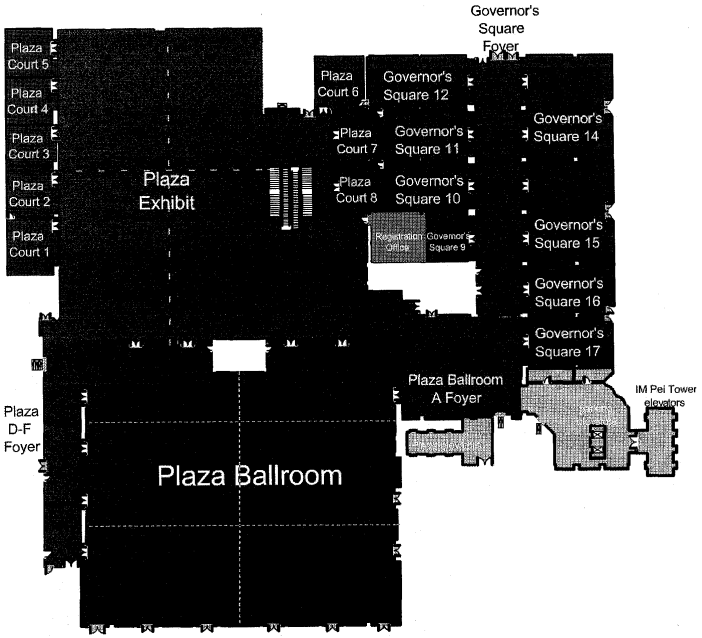


The 14th Conference of the

North American Chapter of the Association for Computational Linguistics:

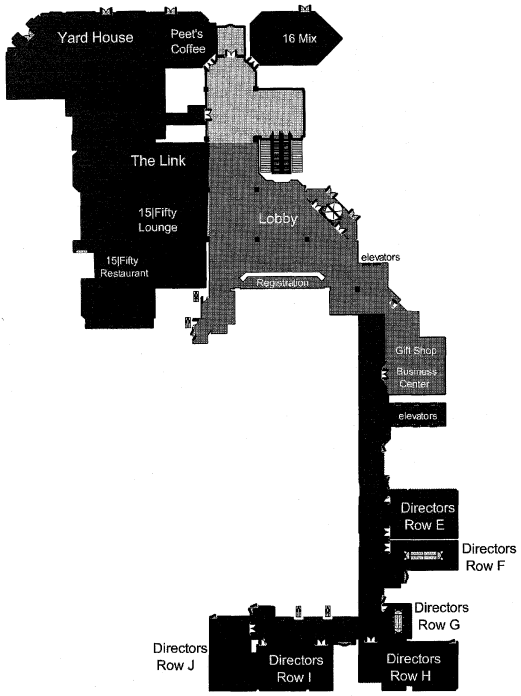
Human Language Technologies

CONFERENCE HANDBOOK



CONCOURSE LEVEL

LOBBY LEVEL



Cover design by Aurelia Bunescu
Handbook assembled by Matt Post
Printing by Omnipress of Madison, Wisconsin

Contents


WHAT WE
ANALYZE
INTERPRET
PRODUCE
MAKE
COMMUNICATE
DESIGN
CREATE
GENERATE
MAKE
DESCRIBE
INTERPRET
IMPLEMENT
REPORT
DEVELOP
FIGURE OUT
DECODE
FOCUS ON
THINK
BUILDS
MOVES
GUIDES
SHAPES
HELPS
FUELS
LIFTS
EXPLORES
MOTIVATES
BREAKS
ENGAGES
INFLUENCES
ADVANCES
DRIVES
SHIFTS
MAXIMIZES
PROMOTES
ENHANCES
COMMUNITIES
MARKETS
ECONOMIES
OPINIONS
CULTURES
WEALTH
ENVIRONMENTS
POSSIBILITIES
OPINIONS
CULTURES
BOUNDARIES
CAREERS
OPPORTUNITIES
IDEAS
TECHNOLOGY
INDUSTRIES
LEADERS
Make your mark.
The world's top investors, traders and leaders depend on our information and news. We need your ideas and passion to help us meet some of the biggest challenges around.
jobs.bloomberg.com
Bloomberg
/bloombergcareers
©2015 Bloomberg L.P. All rights reserved. S571691131 0315

1

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 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies or NAACL HLT 2015 for short.

This year, we received the largest number of submissions in the history of NAACL: a total of 714 submissions with 402 long paper submissions and 312 short papers submissions. From these, 117 long papers (62 oral presentations and 55 poster presentations) and 69 short papers (24 oral presentations and 45 poster presentations) were accepted to appear at the conference.

The submissions to NAACL HLT 2015 were assigned to 18 technical areas including a new topic area called *Language and Vision*. This track was introduced with an intent to broaden research on natural language processing that is situated in a rich visual and perceptual context. We received 16 submissions for this area and seven of them will be presented at the conference.

For NAACL HLT 2015 we initiated a meta review process, where each paper received an analysis of the merits of the paper from the area chair's perspective that was based on the reviewer comments, the reviewer discussion and the author rebuttal. We found the meta reviews very helpful in consolidating the reviews and providing justifications for final decisions. As this was an experiment this year, the meta reviews were not sent to the authors.

Based on comments from reviewers, nominations from area chairs, and rankings from the best paper committee, three papers were selected to receive the best paper awards at the conference.

Continuing the tradition, NAACL HLT 2015 will feature 19 papers which were accepted for publication in the Transactions of the Association for Computational Linguistics (TACL). The TACL papers were split into 10 oral presentations and 9 poster presentatons.

We are very pleased to have two exciting keynote talks: one by Professor Lillian Lee (Cornell University) and the other by Professor Fei-Fei Li (Stanford University).

There are many people to thank for who have worked diligently to make NAACL HLT 2015 possible. Thanks to the 32 area chairs for their hard work on recruiting reviewers, managing reviews, leading discussions, and making recommendations. All the area chairs are listed in the Program Committee section of the Front Matter. Thanks to Chris Callison-Burch, David Mimno, Sameer Pradhan, and Philip Resnik for stepping in to serve as area co-chairs at the last minute when we were faced with an unexpectedly large number of submissions in some tracks.

Following what was done in the last NAACL conference, we used the paper assignment tool developed by Mark Dredze to assign papers to reviewers. Thanks to Mark Dredze and Jiang Guo for their hard work on assigning papers to reviewers based on their preferences. We had to especially rely on this tool this year because the distribution of submissions across areas was very different from past trends.

This program certainly would not be possible without the help of the 460 reviewers. Their names are listed in the Program Committee section. In particular, 116 reviewers from this list were recognized by the area chairs as best reviewers who have turned in exceptionally well-written and constructive reviews and who have actively engaged in the post-rebuttal discussions. The names of the best reviewers are marked with * in the list of reviewers.

We are also indebted to the best paper award committee which consists of Claire Cardie, Daniel Gildea, Daniel Marcu, and Fernando Pereira. Their time and effort in recommending the best paper awards is much appreciated.

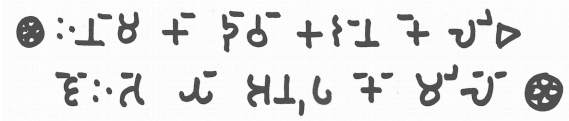
We also would like to thank Hal Daumé III, Kristina Toutanova, and Lucy Vanderwende for generously sharing their experience in organizing prior NAACL/ACL conferences and for their advice. We are grateful for the guidance and the support of the NAACL president Hal Daumé III, and the NAACL board. We also would like to thank the publication co-chairs Matt Post and Adam Lopez for putting together the proceedings and the conference handbook; and Paolo Gai and Rich Gerber from Softconf for always being responsive to our requests.

We would like to thank the ACL Business Manager Priscilla Rasmussen. She was our *go to* person who knew all details of the conference in and out. We are very grateful for her help.

Finally, this conference could not have happened without the efforts of the general chair, Rada

Mihalcea. She made sure the various sections of NAACL organization worked well together. Her monthly newsletters informed all the organizers about what was being done by everyone else. We are very thankful for her leadership in the organization of NAACL HLT 2015.

We hope you will enjoy NAACL HLT 2015!



NAACL HLT 2015 Program Co-Chairs
Joyce Chai, Michigan State University
Anoop Sarkar, Simon Fraser University

Organizing Committee

General Chair
Rada Mihalcea, University of Michigan, USA

Local Arrangements Chair
Priscilla Rasmussen, ACL Business Manager

Program Committee Co-chairs
Joyce Chai, Michigan State University, USA
Anoop Sarkar, Simon Fraser University, Canada

Workshop Co-chairs
Cornelia Caragea, University of North Texas, USA
Bing Liu, University of Illinois at Chicago, USA

Tutorial Co-chairs
Yang Liu, University of Texas at Dallas, USA
Thamar Solorio, University of Houston, USA

Student Research Workshop Co-chairs
Student Co-chairs
Shibamouli Lahiri, University of Michigan, USA
Karen Mazidi, University of North Texas, USA
Alisa Zhila, Instituto Politécnico Nacional (National Polytechnic Institute), Mexico
Faculty Advisors
Diana Inkpen, University of Ottawa, Canada
Smaranda Muresan, Columbia University, USA

Demo Co-chairs
Matt Gerber, University of Virginia, USA
Catherine Havasi, Massachusetts Institute of Technology, USA
Finley Lacatusu, Language Computer Corporation, USA

Student Volunteer Coordinator
Annie Louis, University of Edinburgh, UK

Reviewing Coordinators
Mark Dredze, Johns Hopkins University
Jiang Guo, Harbin Institute of Technology

Local Sponsorship Chair
Kevin B. Cohen, University of Colorado, USA

Publicity Chair
Saif M. Mohammad, National Research Council Canada

Publication Co-chairs
Matt Post, Johns Hopkins University, USA
Adam Lopez, University of Edinburgh, UK

Conference Handbook Editor
Matt Post, Johns Hopkins University

Website Chair
Peter Ljunglöf, University of Gothenburg and Chalmers University of Technology, Sweden

Business Manager
Priscilla Rasmussen

Program Committee

Program Committee Co-chairs
Joyce Chai, Michigan State University, USA
Anoop Sarkar, Simon Fraser University, Canada

Area Chairs

Dialogue and Interactive Systems
Jason D. Williams, Microsoft Research

Discourse and Pragmatics
Ani Nenkova, University of Pennsylvania
Vincent Ng, University of Texas at Dallas

Generation and Summarization
Chris Callison-Burch, University of Pennsylvania
Fei Liu, Carnegie Mellon University

Information Extraction and Question Answering
Alan Ritter, The Ohio State University
Bowen Zhou, IBM Research

Information Retrieval
David Smith, Northeastern University

Language and Vision
Julia Hockenmaier, University of Illinois at Urbana-Champaign

Language Resources and Evaluation
Will Lewis, Microsoft Research
Sameer Pradhan, Harvard University

Linguistic and Psycholinguistic Aspects of CL
William Schuler, The Ohio State University

Machine Learning for NLP
Amarnag Subramanya, Google Research
Tong Zhang, Baidu Inc. and Rutgers University

Machine Translation
Bill Byrne, Cambridge University
Hany Hassan, Microsoft Research
Kevin Knight, Information Sciences Institute, University of Southern California
Haitao Mi, IBM Research

NLP for Web, Social Media and Social Sciences
Brendan O'Connor, University of Massachusetts
Philip Resnik, University of Maryland

NLP-enabled Technology
Richard Socher, Stanford University

Phonology, Morphology, and Word Segmentation
Sami Virpioja, Lingsoft and Aalto University

Semantics
Dipanjan Das, Google
William Dolan, Microsoft Research
Luke Zettlemoyer, University of Washington

Sentiment Analysis and Opinion Mining
Saif M. Mohammad, National Research Council Canada
Jerry Zhu, University of Wisconsin-Madison

Spoken Language Processing

Xiaodong He, Microsoft Research

Tagging, Chunking, Syntax, and Parsing
Noah Smith, Carnegie Mellon University
Yue Zhang, Singapore University of Technology and Design

Text Categorization and Topic Models
Jordan Boyd-Graber, University of Colorado at Boulder
David Mimno, Cornell University

Meal Info

The following meals are provided as part of your registration fee:

- A full buffet breakfast will be provided each day in the Plaza Exhibit (foyer)
- Mid-Morning breaks include coffee and tea in the Plaza Exhibit (foyer)
- Mid-Afternoon breaks include coffee, tea, soda, water, and snacks in the Plaza Exhibit (foyer)
- A full dinner buffer is provided during the poster sessions on Monday and Tuesday evenings in the Plaza Exhibit (foyer)

Lunch is provided for students on Monday, but you are otherwise on your own.

2

Tutorials: Sunday, May 31

Overview

7:30 – 6:00	**Registration**	*Plaza Exhibit All*
7:30 – 9:00	**Breakfast**	*Plaza Exhibit All*
9:00 – 12:30	**Morning Tutorials**	
	Hands-on Learning to Search for Structured Prediction *Hal Daumé III, John Langford, Kai-Wei Chang, He He, and Sudha Rao*	*Governor's Square 15*
	Crowdsourcing for NLP *Chris Callison-Burch, Lyle Ungar, and Ellie Pavlick*	*Governor's Square 17*
	The Logic of AMR: Practical, Unified, Graph-Based Sentence Semantics for NLP *Nathan Schneider, Jeffrey Flanigan, and Tim O'Gorman*	*Governor's Square 16*
10:30 – 11:00	**Coffee break**	
12:30 – 2:00	**Lunch break**	
2:00 – 5:30	**Afternoon Tutorials**	
	Social Media Predictive Analytics *Svitlana Volkova, Benjamin Van Durme, David Yarowsky, and Yoram Bachrach*	*Governor's Square 17*
	Deep Learning and Continuous Representations for Natural Language Processing *Wen-tau Yih, Xiaodong He, and Jianfeng Gao*	*Governor's Square 15*
	Getting the Roles Right: Using FrameNet in NLP *Collin Baker, Nathan Schneider, Miriam R L Petruck, and Michael Ellsworth*	*Governor's Square 16*
3:30 – 4:00	**Coffee break**	
6:00 – 9:00	**Welcome Reception**	*Foyer*

Tutorial 1

Hands-on Learning to Search for Structured Prediction

Hal Daumé III, John Langford, Kai-Wei Chang, He He, and Sudha Rao

Sunday, May 31, 2015, 9:00–12:30pm

Governor's Square 15

Many problems in natural language processing involve building outputs that are structured. The predominant approach to structured prediction is "global models" (such as conditional random fields), which have the advantage of clean underlying semantics at the cost of computational burdens and extreme difficulty in implementation. An alternative strategy is the "learning to search" (L2S) paradigm, in which the structured prediction task is cast as a sequential decision making process. One can then devise training-time algorithms that learn to make near optimal collective decisions. This paradigm has been gaining increasing traction over the past five years: most notably in dependency parsing (e.g., MaltParser, ClearNLP, etc.), but also much more broadly in less "sequential" tasks like entity/relation classification and even graph prediction problems found in social network analysis and computer vision.

Hal Daumé III is an associate professor in Computer Science at the University of Maryland, College Park. He holds joint appointments in UMIACS and Linguistics. His primary research interest is in developing new learning algorithms for prototypical problems that arise in the context of language processing and artificial intelligence. This includes topics like structured prediction, domain adaptation and unsupervised learning; as well as multilingual modeling and affect analysis.

John Langford is a machine learning research scientist. He is the author of the blog hunch.net and the principal developer of Vowpal Wabbit. John works at Microsoft Research New York, of which he was one of the founding members, and was previously affiliated with Yahoo! Research, Toyota Technological Institute, and IBM's Watson Research Center. He studied Physics and Computer Science at the California Institute of Technology, earning a double bachelor's degree in 1997, and received his Ph.D. in Computer Science from Carnegie Mellon University in 2002. He was the program co-chair for the 2012 International Conference on Machine Learning.

Kai-Wei Chang is a doctoral candidate in Computer Science at the University of Illinois at Urbana-Champaign. His research interests lie in designing practical machine learning techniques for large and complex data and applying them to real world applications. He has been working on various topics in Machine learning and Natural Language Processing, including large-scale learning, structured learning, coreference resolution, and relation extraction.

He He is a doctoral student in Computer Science at the University of Maryland, College Park. Her research interests focus on developing fast inference methods for structured prediction problems in machine learning and natural language processing, including features selection, dependency parsing and machine translation.

Sudha Rao is a doctoral student in Computer Science at the University of Maryland, College Park. Her research interests currently lie in exploring how semantic representations like Abstract Meaning Representation can be used in Machine Translation.

This tutorial has precisely one goal: an attendee should leave the tutorial with hands on experience writing small programs to perform structured prediction for a variety of tasks, like sequence labeling, dependency parsing and, time-permitting, more.

This proposed tutorial is unique (to our knowledge) among ACL tutorials in this regard: half of the time spent will be in the style of a "flipped classroom" in which attendees get hands on experience writing structured predictors on their own or in small groups. All course materials (software, exercises, hints, solutions, etc., will be made available at least two weeks prior to the event so that students can download the required data ahead of time; we will also bring copies on USB in case there is a problem with the internet).

The first half of the tutorial will be mostly "lecture" style, in which we will cover the basics of how learning to search works for structured prediction. The goal is to provide enough background information that students can understand how to write and debug their own predictors, but the emphasis will not be on how to build new machine learning algorithms. This will also include a brief tutorial on the basics of Vowpal Wabbit, to the extent necessary to understand its structured prediction interface. The second half of the tutorial will focus on hands-on exploration of structured prediction using the Vowpal Wabbit python "learning to search" interface; a preliminary python notebook explaining the interface can be viewed at `http://tinyurl.com/pyvwsearch`; an elaborated version of this notebook will serve as the backbone for the "hands on" part of the tutorial, paired with exercises.

Tutorial 2

Crowdsourcing for NLP

Chris Callison-Burch, Lyle Ungar, and Ellie Pavlick

Sunday, May 31, 2015, 9:00–12:30pm

Governor's Square 17

Crowdsourced applications to scientific problems is a hot research area, with over 10,000 publications in the past five years. Platforms such as Amazon's Mechanical Turk and CrowdFlower provide researchers with easy access to large numbers of workers. The crowd's vast supply of inexpensive, intelligent labor allows people to attack problems that were previously impractical and gives potential for detailed scientific inquiry of social, psychological, economic, and linguistic phenomena via massive sample sizes of human annotated data. We introduce crowdsourcing and describe how it is being used in both industry and academia. Crowdsourcing is valuable to computational linguists both (a) as a source of labeled training data for use in machine learning

---

Chris Callison-Burch is the Aravind K Joshi term assistant professor in the Computer and Information Science Department at the University of Pennsylvania. Before joining Penn, he was a research faculty member at the Center for Language and Speech Processing at Johns Hopkins University for 6 years. He was the Chair of the Executive Board of the North American chapter of the Association for Computational Linguistics (NAACL) from 2011-2013, and he has served on the editorial boards of the journals Transactions of the ACL (TACL) and Computational Linguistics. He is a Sloan Research Fellow, and he has received faculty research awards from Google, Microsoft and Facebook in addition to funding from DARPA and the NSF. Chris teaches a semester-long course on Crowdsourcing at Penn (`http://crowdsourcing-class.org/`).

Lyle Ungar is a Professor of Computer and Information Science at the University of Pennsylvania. He also holds appointments in several other departments in the Engineering, Medicine, and Business Schools. Dr. Ungar received a B.S. from Stanford University and a Ph.D. from M.I.T. He has published over 200 articles and is co-inventor on eight patents. His current research includes machine learning, data mining, and text mining, and uses social media to better understand the drivers of physical and mental well-being. Lyle's research group collects MTurk crowdsourced labels on natural language data such Facebook posts and tweets, which they use for a variety of NLP and psychology studies. Lyle (with collaborators) has given highly successful tutorials on information extraction, sentiment analysis, and spectral methods for NLP at conferences including NAACL, KDD, SIGIR, ICWSM, CIKM, and AAAI. He and his student gave a tutorial on crowdsourcing last year at the Joint Statistical Meetings (JSM).

Ellie Pavlick is a Ph.D. student at the University of Pennsylvania. Ellie received her B.A. in economics from the Johns Hopkins University, where she began working with Dr. Chris Callison-Burch on using crowdsourcing to create low-cost training data for statistical machine translation by hiring non-professional translators and post-editors. Her current research interests include entailment and paraphrase recognition, for which she has looked at using MTurk to provide more difficult linguistic annotations such as discriminating between fine-grained lexical entailment relations and identifying missing lexical triggers in FrameNet. Ellie TAed and helped design the curriculum for the Crowdsourcing and Human Computation course at Penn.

and (b) as a means of collecting computational social science data that link language use to underlying beliefs and behavior. We present case studies for both categories: (a) collecting labeled data for use in natural language processing tasks such as word sense disambiguation and machine translation and (b) collecting experimental data in the context of psychology; e.g. finding how word use varies with age, sex, personality, health, and happiness.

We will also cover tools and techniques for crowdsourcing. Effectively collecting crowdsourced data requires careful attention to the collection process, through selection of appropriately qualified workers, giving clear instructions that are understandable to non-experts, and performing quality control on the results to eliminate spammers who complete tasks randomly or carelessly in order to collect the small financial reward. We will introduce different crowdsourcing platforms, review privacy and institutional review board issues, and provide rules of thumb for cost and time estimates. Crowdsourced data also has a particular structure that raises issues in statistical analysis; we describe some of the key methods to address these issues.

Tutorial 3

The Logic of AMR: Practical, Unified, Graph-Based Sentence Semantics for NLP

Nathan Schneider, Jeffrey Flanigan, and Tim O'Gorman

Sunday, May 31, 2015, 9:00–12:30pm

Governor's Square 16

The Abstract Meaning Representation formalism is rapidly emerging as an important practical form of structured sentence semantics which, thanks to the availability of large-scale annotated corpora, has potential as a convergence point for NLP research. This tutorial unmasks the design philosophy, data creation process, and existing algorithms for AMR semantics. It is intended for anyone interested in working with AMR data, including parsing text into AMRs, generating text from AMRs, and applying AMRs to tasks such as machine translation and summarization.

The goals of this tutorial are twofold. First, it will describe the nature and design principles behind the representation, and demonstrate that it can be practical for annotation. In Part I: The AMR Formalism, participants will be coached in the basics of annotation so that, when working with AMR data in the future, they will appreciate the benefits and limitations of the process by which it was created. Second, the tutorial will survey the state of the art for computation with AMRs. Part II: Algorithms and Applications will focus on the task of parsing English text into AMR graphs, which requires algorithms for alignment, for structured prediction, and for statistical learning. The tutorial will also address graph grammar formalisms that have been recently developed, and future applications such as AMR-based machine translation and summarization.

Nathan Schneider is an annotation schemer and computational modeler for natural language. He has been involved in the design of the AMR formalism since 2012, when he interned with Kevin Knight at ISI. His 2014 dissertation introduced a coarse-grained representation for lexical semantics that facilitates rapid annotation and is practical for broad-coverage statistical NLP. He has also worked on semantic parsing for the FrameNet representation and other forms of syntactic/semantic annotation and processing for social media text. For most of these projects, he led the design of the annotation scheme, guidelines, and workflows, and the training and supervision of annotators.

Tim O'Gorman is a third year Linguistics Ph.D. student at the University of Colorado – Boulder, working with Martha Palmer. He manages CU-Boulder's AMR annotation team, and participated in the Fred Jelinek Memorial Workshop in Prague in 2014, working on mapping the Prague tectogrammatical layer to AMRs. His research areas include implicit arguments, semantic role projection, and linking sentence-level semantics to discourse.

Jeffrey Flanigan is a fifth year Ph.D. candidate at Carnegie Mellon University. He and his collaborators at CMU built the first broad-coverage AMR parser. He also participated in the Fred Jelinek Memorial Workshop in Prague in 2014, working on cross-lingual parsing and generation from AMR. His research areas include machine translation, generation, summarization, and semantic parsing.

Tutorial 4

Social Media Predictive Analytics

Svitlana Volkova, Benjamin Van Durme, David Yarowsky, and Yoram Bachrach

Sunday, May 31, 2015, 2:00–5:30pm

Governor's Square 17

The recent explosion of social media services like Twitter, Google+ and Facebook has led to an interest in social media predictive analytics – automatically inferring hidden information from the large amounts of freely available content. It has a number of applications, including: online targeted advertising, personalized marketing, large-scale passive polling and real-time live polling, personalized recommendation systems and search, and real-time healthcare analytics etc.

In this tutorial, we will describe how to build a variety of social media predictive analytics for inferring latent user properties from a Twitter network including demographic traits, personality, interests, emotions and opinions etc. Our methods will address several important aspects of social media such as: dynamic, streaming nature of the data, multi-relationality in social networks, data collection and annotation biases, data and model sharing, generalization of the existing models, data drift, and scalability to other languages.

We will start with an overview of the existing approaches for social media predictive analytics. We will describe the state-of-the-art static (batch) models and features. We will then present models for streaming (online) inference from single and multiple data streams; and formulate a latent attribute prediction task as a sequence-labeling problem. Finally, we present several techniques for dynamic (iterative) learning and prediction using active learning setup with rationale annotation and filtering. The tutorial will conclude with a practice session focusing on

---

Svitlana Volkova is a Ph.D. Candidate in Computer Science at the Center for Language and Speech Processing, Johns Hopkins University. She works on machine learning and natural language processing techniques for social media predictive analytics. She develops batch and streaming (dynamic) models for automatically inferring psycho-demographic profiles from social media data streams, fine-grained emotion detection and sentiment analysis for under-explored languages and dialects in microblogs, effective interactive and iterative rationale annotation via crowdsourcing.

Benjamin Van Durme is the Chief Lead of Text Research at the Human Language Technology Center of Excellence, and an Assistant Research Professor at the Center for Language and Speech Processing. He works on natural language processing (specifically computational semantics), predictive analytics in social media and streaming/randomized algorithms.

David Yarowsky is a Professor at the Center for Language and Speech Processing, Johns Hopkins University. His research interests include natural language processing and spoken language systems, machine translation, information retrieval, very large text databases and machine learning. His research focuses on word sense disambiguation, minimally supervised induction algorithms in NLP, and multilingual natural language processing.

Yoram Bachrach is a researcher in the Online Services and Advertising group at Microsoft Research Cambridge UK. His research area is artificial intelligence (AI), focusing on multi-agent systems and computational game theory. Computational game theory combines the theoretical foundations of economics and game theory with creative solutions from AI and computer science.

walk-through examples for predicting latent user properties e.g., political preferences, income, education level, life satisfaction and emotions emanating from user communications on Twitter.

Tutorial 5

Deep Learning and Continuous Representations for Natural Language Processing

Wen-tau Yih, Xiaodong He, and Jianfeng Gao

Sunday, May 31, 2015, 2:00–5:30pm

Scott Wen-tau Yih is a Researcher in the Machine Learning Group at Microsoft Research Redmond. His research interests include natural language processing, machine learning and information retrieval. Yih received his Ph.D. in computer science at the University of Illinois at Urbana-Champaign. His work on joint inference using integer linear programming (ILP) [Roth & Yih, 2004] helped the UIUC team win the CoNLL-05 shared task on semantic role labeling, and the approach has been widely adopted in the NLP community. After joining MSR in 2005, he has worked on email spam filtering, keyword extraction and search & ad relevance. His recent work focuses on continuous semantic representations using neural networks and matrix/tensor decomposition methods, with applications in lexical semantics, knowledge base embedding and question answering. Yih received the best paper award from CoNLL-2011 and has served as area chairs (HLT-NAACL-12, ACL-14) and program co-chairs (CEAS-09, CoNLL-14) in recent years.

Xiaodong He is a Researcher of Microsoft Research, Redmond, WA, USA. He is also an Affiliate Professor in Electrical Engineering at the University of Washington, Seattle, WA, USA. His research interests include deep learning, information retrieval, natural language understanding, machine translation, and speech recognition. Dr. He has published a book and more than 70 technical papers in these areas, and has given tutorials at international conferences in these fields. In benchmark evaluations, he and his colleagues have developed entries that obtained No. 1 place in the 2008 NIST Machine Translation Evaluation (NIST MT) and the 2011 International Workshop on Spoken Language Translation Evaluation (IWSLT), both in Chinese-English translation, respectively. He serves as Associate Editor of IEEE Signal Processing Magazine and IEEE Signal Processing Letters, as Guest Editors of IEEE TASLP for the Special Issue on Continuous-space and related methods in natural language processing, and Area Chair of NAACL2015. He also served as GE for several IEEE Journals, and served in organizing committees and program committees of major speech and language processing conferences in the past. He is a senior member of IEEE and a member of ACL.

Jianfeng Gao is a Principal Researcher of Microsoft Research, Redmond, WA, USA. His research interests include Web search and information retrieval, natural language processing and statistical machine learning. Dr. Gao is the primary contributor of several key modeling technologies that help significantly boost the relevance of the Bing search engine. His research has also been applied to other MS products including Windows, Office and Ads. In benchmark evaluations, he and his colleagues have developed entries that obtained No. 1 place in the 2008 NIST Machine Translation Evaluation in Chinese-English translation. He was Associate Editor of ACM Trans on Asian Language Information Processing, (2007 to 2010), and was Member of the editorial board of Computational Linguistics (2006 – 2008). He also served as area chairs for ACL-IJCNLP2015, SIGIR2015, SIGIR2014, IJCAI2013, ACL2012, EMNLP2010, ACL-IJCNLP 2009, etc. Dr. Gao recently joined Deep Learning Technology Center at MSR-NExT, working on Enterprise Intelligence.

Governor's Square 15

Deep learning techniques have demonstrated tremendous success in the speech and language processing community in recent years, establishing new state-of-the-art performance in speech recognition, language modeling, and have shown great potential for many other natural language processing tasks. The focus of this tutorial is to provide an extensive overview on recent deep learning approaches to problems in language or text processing, with particular emphasis on important real-world applications including language understanding, semantic representation modeling, question answering and semantic parsing, etc.

In this tutorial, we will first survey the latest deep learning technology, presenting both theoretical and practical perspectives that are most relevant to our topic. We plan to cover common methods of deep neural networks and more advanced methods of recurrent, recursive, stacking and convolutional networks. In addition, we will introduce recently proposed continuous-space representations for both semantic word embedding and knowledge base embedding, which are modeled by either matrix/tensor decomposition or neural networks.

Next, we will review general problems and tasks in text/language processing, and underline the distinct properties that differentiate language processing from other tasks such as speech and image object recognition. More importantly, we highlight the general issues of natural language processing, and elaborate on how new deep learning technologies are proposed and fundamentally address these issues. We then place particular emphasis on several important applications, including (1) machine translation, (2) semantic information retrieval and (3) semantic parsing and question answering. For each task, we will discuss what particular architectures of deep learning models are suitable given the nature of the task, and how learning can be performed efficiently and effectively using end-to-end optimization strategies.

Tutorial 6

Getting the Roles Right: Using FrameNet in NLP

Collin Baker, Nathan Schneider, Miriam R L Petruck, and Michael Ellsworth

Sunday, May 31, 2015, 2:00–5:30pm

Governor's Square 16

The FrameNet lexical database (Fillmore & Baker 2010, Ruppenhofer et al. 2006, `http://framenet.icsi.berkeley.edu`), covers roughly 13,000 lexical units (word senses) for the core Engish lexicon, associating them with roughly 1,200 fully defined semantic frames; these frames and their roles cover the majority of event types in everyday, non-specialist text, and they are documented with 200,000 manually annotated examples. This tutorial will teach attendees what they need to know to start using the FrameNet lexical database as part of an NLP system. We will cover the basics of Frame Semantics, explain how the database was created, introduce the Python API and the state of the art in automatic frame semantic role labeling systems; and we will discuss FrameNet collaboration with commercial partners. Time permitting, we will present new research on frames and annotation of locative relations, as well as corresponding metaphorical uses, along with information about how frame semantic roles can aid the interpretation of metaphors.

Collin Baker has been Project Manager of the FrameNet Project since 2000. His research interests include FrameNets in other languages (Loenneker-Rodman & Baker 2009), aligning FrameNet to other lexical resources (Fellbaum & Baker 2013, Ferrandez et al 2010), linking to ontologies and reasoning (Scheffczyk et al. 2010), and the frame semantics of metaphor.

Nathan Schneider has worked on a coarse-grained representation for lexical semantics (2014 dissertation at Carnegie Mellon University) and the design of the Abstract Meaning Representation (AMR; Banarescu et al. 2014). Nathan helped develop the leading open-source frame-semantic parser for English, SEMAFOR (Das et al. 2010, 2014) (`http://demo.ark.cs.cmu.edu/parse`), as well as a Python interface to the FrameNet lexicon (with Chuck Wooters) that is part of the NLTK suite.

Miriam R. L. Petruck 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.

Michael Ellsworth has been involved with FrameNet for well over a decade. His chief focus is on semantic relations in FrameNet (Ruppenhofer et al. 2006), how they can be used for paraphrase (Ellsworth & Janin 2007), and mapping to other resources (Scheffczyk et al. 2006, Ferrandez et al. 2010). Increasingly, he has examined the connection of FrameNet to syntax and the Constructicon (Torrent & Ellsworth 2013, Ziem & Ellsworth 2015), including in his pending dissertation on the constructions and frame semantics of emotion.

Welcome Reception

Sunday, May 31, 2015, 6:00pm – 9:00pm

Sheraton Denver Downtown Hotel (conference venue)
Foyer

Catch up with your colleagues at the **Welcome Reception**! It will be held immediately following the Tutorials on Sunday, May 31 at 6:00pm in the Plaza Exhibit (foyer) of the Sheraton Denver Downtown Hotel (the conference venue). Refreshments and a light dinner will be provided, and a cash bar will be available.

3

Main Conference: Monday, June 1

Overview

7:30 – 8:45	**Registration and Breakfast**	*Plaza Exhibit All*
8:45 – 9:00	**Welcome to NAACL HLT 2015**	*Plaza Ballroom A, B, & C*
9:00 – 10:10	**Invited Talk: "Big data pragmatics!", or, "Putting the ACL in computational social science", or, if you think these title alternatives could turn people on, turn people off, or otherwise have an effect, this talk might be for you (Lillian Lee)**	*Plaza Ballroom A, B, & C*
10:10 – 10:40	**Break**	*Plaza Exhibit All*

Session 1

10:40 – 12:20	Semantics (Long Papers)	Tagging, Chunking, Syntax and Parsing (Long + TACL Papers)	Information Retrieval, Text Categorization, Topic Modeling (Long Papers)
	Plaza Ballroom A & B	*Plaza Ballroom D & E*	*Plaza Ballroom F*

12:20 – 2:00	**Lunch**	

Session 2

2:00 – 3:15	Generation and Summarization (Long Papers)	Language and Vision (Long Papers)	NLP for Web, Social Media and Social Sciences (Long Papers)
	Plaza Ballroom A & B	*Plaza Ballroom D & E*	*Plaza Ballroom F*

Session 3

3:15 – 4:00	Generation and Summarization (Short Papers)	Information Extraction and Question Answering (Short Papers)	Machine Learning for NLP (Short Papers)
	Plaza Ballroom A & B	*Plaza Ballroom D & E*	*Plaza Ballroom F*

4:00 – 4:30	**Break**	*Plaza Exhibit All*
4:30 – 6:00	**One minute madness (Long + TACL papers)**	*Plaza Ballroom A, B, & C*
6:00 – 9:00	**SRW Poster Session**	*Plaza Ballroom A, B, & C*
6:00 – 7:30	**Poster session 1A: Long + TACL papers**	*Plaza Ballroom A, B, & C*
7:30 – 9:00	**Poster session 1B: Long + TACL papers**	*Plaza Ballroom A, B, & C*

Keynote Address: Lillian Lee

"Big data pragmatics!", or, "Putting the ACL in computational social science", or, if you think these title alternatives could turn people on, turn people off, or otherwise have an effect, this talk might be for you

Monday, June 1, 2015, 9:00–10:10am

Plaza Ballroom A, B, & C

Abstract: What effect does language have on people?

You might say in response, "Who are you to discuss this problem?" and you would be right to do so; this is a Major Question that science has been tackling for many years. But as a field, I think natural language processing and computational linguistics have much to contribute to the conversation, and I hope to encourage the community to further address these issues.

This talk will focus on the effect of phrasing, emphasizing aspects that go beyond just the selection of one particular word over another. The issues we'll consider include: Does the way in which something is worded in and of itself have an effect on whether it is remembered or attracts attention, beyond its content or context? Can we characterize how different sides in a debate frame their arguments, in a way that goes beyond specific lexical choice (e.g., "pro-choice" vs. "pro-life")? The settings we'll explore range from movie quotes that achieve cultural prominence; to posts on Facebook, Wikipedia, Twitter, and the arXiv; to framing in public discourse on the inclusion of genetically-modified organisms in food.

Joint work with Lars Backstrom, Justin Cheng, Eunsol Choi, Cristian Danescu-Niculescu-Mizil, Jon Kleinberg, Bo Pang, Jennifer Spindel, and Chenhao Tan.

Biography: Lillian Lee is a professor of computer science and of information science at Cornell University, and the co-Editor-in-Chief, together with Michael Collins, of Transactions of the ACL. Her research interests include natural language processing and computational social science. She is the recipient of the inaugural Best Paper Award at HLT-NAACL 2004 (joint with Regina Barzilay), a citation in "Top Picks: Technology Research Advances of 2004" by Technology Research News (also joint with Regina Barzilay), and an Alfred P. Sloan Research Fellowship; and in 2013, she was named a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI). Her group's work has received several mentions in the popular press, including The New York Times, NPR's All Things Considered, and NBC's The Today Show, and one of her co-authored papers was publicly called "boring" by Youtubers Rhett and Link, in a video viewed over 1.8 million times.

Session 1 Overview – Monday, June 1, 2015

Track A	**Track B**	**Track C**	
Semantics (Long Papers) Plaza Ballroom A & B	*Tagging, Chunking, Syntax and Parsing (Long + TACL Papers)* Plaza Ballroom D & E	*Information Retrieval, Text Categorization, Topic Modeling (Long Papers)* Plaza Ballroom F	
Unsupervised Induction of Semantic Roles within a Reconstruction-Error Minimization Framework *Titov and Khoddam*	Randomized Greedy Inference for Joint Segmentation, POS Tagging and Dependency Parsing *Zhang, Li, Barzilay, and Darwish*	A Hybrid Generative/Discriminative Approach To Citation Prediction *Tanner and Charniak*	10:40
Predicate Argument Alignment using a Global Coherence Model *Wolfe, Dredze, and Van Durme*	An Incremental Algorithm for Transition-based CCG Parsing *Ambati, Deoskar, Johnson, and Steedman*	Weakly Supervised Slot Tagging with Partially Labeled Sequences from Web Search Click Logs *Kim, Jeong, Stratos, and Sarikaya*	11:05
Improving unsupervised vector-space thematic fit evaluation via role-filler prototype clustering *Greenberg, Sayeed, and Demberg*	Because Syntax Does Matter: Improving Predicate-Argument Structures Parsing with Syntactic Features *Ribeyre, Villemonte de la Clergerie, and Seddah*	Not All Character N-grams Are Created Equal: A Study in Authorship Attribution *Sapkota, Bethard, Montes, and Solorio*	11:30
A Compositional and Interpretable Semantic Space *Fyshe, Wehbe, Talukdar, Murphy, and Mitchell*	[TACL] Exploring Compositional Architectures and Word Vector Representations for Prepositional Phrase Attachment *Belinkov, Lei, Barzilay, and Globerson*	Effective Use of Word Order for Text Categorization with Convolutional Neural Networks *Johnson and Zhang*	11:55

Parallel Session 1

Session 1A: Semantics (Long Papers)

Plaza Ballroom A & B *Chair: Hoifung Poon*

Unsupervised Induction of Semantic Roles within a Reconstruction-Error Minimization Framework
Ivan Titov and Ehsan Khoddam 10:40–11:05

We introduce a new approach to unsupervised estimation of feature-rich semantic role labeling models. Our model consists of two components: (1) an encoding component: a semantic role labeling model which predicts roles given a rich set of syntactic and lexical features; (2) a reconstruction component: a tensor factorization model which relies on roles to predict argument fillers. When the components are estimated jointly to minimize errors in argument reconstruction, the induced roles largely correspond to roles defined in annotated resources. Our method performs on par with most accurate role induction methods on English and German, even though, unlike these previous approaches, we do not incorporate any prior linguistic knowledge about the languages.

Predicate Argument Alignment using a Global Coherence Model
Travis Wolfe, Mark Dredze, and Benjamin Van Durme 11:05–11:30

We present a joint model for predicate argument alignment. We leverage multiple sources of semantic information, including temporal ordering constraints between events. These are combined in a max-margin framework to find a globally consistent view of entities and events across multiple documents, which leads to improvements on the state-of-the-art.

Improving unsupervised vector-space thematic fit evaluation via role-filler prototype clustering
Clayton Greenberg, Asad Sayeed, and Vera Demberg 11:30–11:55

Most recent unsupervised methods in vector space semantics for assessing thematic fit (e.g. Erk, 2007; Baroni and Lenci, 2010; Sayeed and Demberg, 2014) create prototypical role-fillers without performing word sense disambiguation. This leads to a kind of sparsity problem: candidate role-fillers for different senses of the verb end up being measured by the same "yardstick", the single prototypical role-filler. In this work, we use three different feature spaces to construct robust unsupervised models of distributional semantics. We show that correlation with human judgements on thematic fit estimates can be improved consistently by clustering typical role-fillers and then calculating similarities of candidate role-fillers with these cluster centroids. The suggested methods can be used in any vector space model that constructs a prototype vector from a non-trivial set of typical vectors.

A Compositional and Interpretable Semantic Space
Alona Fyshe, Leila Wehbe, Partha P. Talukdar, Brian Murphy, and Tom M. Mitchell 11:55–12:20

Vector Space Models (VSMs) of Semantics are useful tools for exploring the semantics of single words, and the composition of words to make phrasal meaning. While many methods can estimate the meaning (i.e. vector) of a phrase, few do so in an interpretable way. We introduce a new method (CNNSE) that allows word and phrase vectors to adapt to the notion of composition. Our method learns a VSM that is both tailored to support a chosen semantic composition operation, and whose resulting features have an intuitive interpretation. Interpretability allows for the exploration of phrasal semantics, which we leverage to analyze performance on a behavioral task.

Session 1B: Tagging, Chunking, Syntax and Parsing (Long + TACL Papers)

Plaza Ballroom D & E *Chair: Noah A. Smith*

Randomized Greedy Inference for Joint Segmentation, POS Tagging and Dependency Parsing
Yuan Zhang, Chengtao Li, Regina Barzilay, and Kareem Darwish 10:40–11:05

In this paper, we introduce a new approach for joint segmentation, POS tagging and dependency parsing. While joint modeling of these tasks addresses the issue of error propagation inherent in traditional pipeline architectures, it also complicates the inference task. Past research has addressed this challenge by placing constraints on the scoring function. In contrast, we propose an approach that can handle arbitrarily complex scoring functions. Specifically, we employ a randomized greedy algorithm that jointly predicts segmentations, POS tags and dependency trees. Moreover, this architecture readily handles different segmentation tasks, such as morphological segmentation for Arabic and word segmentation for Chinese. The joint model outperforms the state-of-the-art systems on three datasets, obtaining 2.1% TedEval absolute gain against the best published results in the 2013 SPMRL shared task.

An Incremental Algorithm for Transition-based CCG Parsing
Bharat Ram Ambati, Tejaswini Deoskar, Mark Johnson, and Mark Steedman 11:05–11:30

Incremental parsers have potential advantages for applications like language modeling for machine translation and speech recognition. We describe a new algorithm for incremental transition-based Combinatory Categorial Grammar parsing. As English CCGbank derivations are mostly right branching and non-incremental, we design our algorithm based on the dependencies resolved rather than the derivation. We introduce two new actions in the shift-reduce paradigm based on the idea of 'revealing' (Pareschi and Steedman, 1987) the required information during parsing. On the standard CCGbank test data, our algorithm achieved improvements of 0.88% in labeled and 2.0% in unlabeled F-score over a greedy non-incremental shift-reduce parser.

Because Syntax Does Matter: Improving Predicate-Argument Structures Parsing with Syntactic Features
Corentin Ribeyre, Eric Villemonte de la Clergerie, and Djamé Seddah 11:30–11:55

Parsing full-fledged predicate-argument structures in a deep syntax framework requires graphs to be predicted. Using the DeepBank (Flickinger et al., 2012) and the Predicate-Argument Structure treebank (Miyao and Tsujii, 2005) as a test field, we show how transition-based parsers, extended to handle connected graphs, benefit from the use of topologically different syntactic features such as dependencies, tree fragments, spines or syntactic paths, bringing a much needed context to the parsing models, improving notably over long distance dependencies and elided coordinate structures. By confirming this positive impact on an accurate 2nd-order graph-based parser (Martins and Almeida, 2014), we establish a new state-of-the-art on these data sets.

[TACL] Exploring Compositional Architectures and Word Vector Representations for Prepositional Phrase Attachment
Yonatan Belinkov, Tao Lei, Regina Barzilay, and Amir Globerson 11:55–12:20

Prepositional phrase (PP) attachment disambiguation is a known challenge in syntactic parsing. The lexical sparsity associated with PP attachments motivates research in word representations that can capture pertinent syntactic and semantic features of the word. One promising solution is to use word vectors induced from large amounts of raw text. However, state-of-the-art systems that employ such representations yield modest gains in PP attachment accuracy. In this paper, we show that word vector representations can yield significant PP attachment performance gains. This is achieved via a non-linear architecture that is discriminatively trained to maximize PP attachment accuracy. The architecture is initialized with word vectors trained from unlabeled data, and relearns those to maximize attachment accuracy. We obtain additional performance gains with alternative representations such as dependency-based word vectors. When tested on both English and Arabic datasets, our method outperforms both a strong SVM classifier and state-of-the-art parsers. For instance, we achieve 82.6% PP attachment accuracy on Arabic, while the Turbo and Charniak self-trained parsers obtain 76.7% and 80.8% respectively.

Session 1C: Information Retrieval, Text Categorization, Topic Modeling (Long Papers)

Plaza Ballroom F *Chair: Jordan Boyd-Graber*

A Hybrid Generative/Discriminative Approach To Citation Prediction

Chris Tanner and Eugene Charniak 10:40–11:05

Text documents of varying nature (e.g., summary documents written by analysts or published, scientific papers) often cite others as a means of providing evidence to support a claim, attributing credit, or referring the reader to related work. We address the problem of predicting a document's cited sources by introducing a novel, discriminative approach which combines a content-based generative model (LDA) with author-based features. Further, our classifier is able to learn the importance and quality of each topic within our corpus – which can be useful beyond this task – and preliminary results suggest its metric is competitive with other standard metrics (Topic Coherence). Our flagship system, Logit-Expanded, provides state-of-the-art performance on the largest corpus ever used for this task.

Weakly Supervised Slot Tagging with Partially Labeled Sequences from Web Search Click Logs

Young-Bum Kim, Minwoo Jeong, Karl Stratos, and Ruhi Sarikaya 11:05–11:30

In this paper, we apply a weakly-supervised learning approach for slot tagging using con- ditional random fields by exploiting web search click logs. We extend the constrained lattice training of Täckström et al. (2013) to non-linear conditional random fields in which latent variables mediate between observations and labels. When combined with a novel initialization scheme that leverages unlabeled data, we show that our method gives signifi- cant improvement over strong supervised and weakly-supervised baselines.

Not All Character N-grams Are Created Equal: A Study in Authorship Attribution

Upendra Sapkota, Steven Bethard, Manuel Montes, and Thamar Solorio 11:30–11:55

Character n-grams have been identified as the most successful feature in both single-domain and cross-domain Authorship Attribution (AA), but the reasons for their discriminative value were not fully understood. We identify subgroups of character n-grams that correspond to linguistic aspects commonly claimed to be covered by these features: morpho-syntax, thematic content and style. We evaluate the predictiveness of each of these groups in two AA settings: a single domain setting and a cross-domain setting where multiple topics are present. We demonstrate that character n-grams that capture information about affixes and punctuation account for almost all of the power of character n-grams as features. Our study contributes new insights into the use of n-grams for future AA work and other classification tasks.

Effective Use of Word Order for Text Categorization with Convolutional Neural Networks

Rie Johnson and Tong Zhang 11:55–12:20

Convolutional neural network (CNN) is a neural network that can make use of the internal structure of data such as the 2D structure of image data. This paper studies CNN on text categorization to exploit the 1D structure (namely, word order) of text data for accurate prediction. Instead of using low-dimensional word vectors as input as is often done, we directly apply CNN to high-dimensional text data, which leads to directly learning embedding of small text regions for use in classification. In addition to a straightforward adaptation of CNN from image to text, a simple but new variation which employs bag-of-word conversion in the convolution layer is proposed. An extension to combine multiple convolution layers is also explored for higher accuracy. The experiments demonstrate the effectiveness of our approach in comparison with state-of-the-art methods.

Student Lunch

Monday, June 1, 2015, 12:30–2:00pm

Plaza Ballroom C

Join your fellow students for a free students-only lunch on Monday, June 1 at 12:30 in the Plaza Ballroom C at the hotel. Entrance tickets will be in your conference badges. This is a chance to get to know others who share similar interests and goals and who may become your lifelong colleagues.

Session 2 Overview – Monday, June 1, 2015

	Track A	Track B	Track C
	Generation and Summarization (Long Papers) Plaza Ballroom A & B	*Language and Vision (Long Papers)* Plaza Ballroom D & E	*NLP for Web, Social Media and Social Sciences (Long Papers)* Plaza Ballroom F
2:00	Transition-Based Syntactic Linearization *Liu, Zhang, Che, and Qin*	What's Cookin'? Interpreting Cooking Videos using Text, Speech and Vision *Malmaud, Huang, Rathod, Johnston, Rabinovich, and Murphy*	TopicCheck: Interactive Alignment for Assessing Topic Model Stability *Chuang, Roberts, Stewart, Weiss, Tingley, Grimmer, and Heer*
2:25	Extractive Summarisation Based on Keyword Profile and Language Model *Xu, Martin, and Mahidadia*	Combining Language and Vision with a Multimodal Skip-gram Model *Lazaridou, Pham, and Baroni*	Inferring latent attributes of Twitter users with label regularization *Mohammady Ardehaly and Culotta*
2:50	HEADS: Headline Generation as Sequence Prediction Using an Abstract Feature-Rich Space *Colmenares, Litvak, Mantrach, and Silvestri*	Discriminative Unsupervised Alignment of Natural Language Instructions with Corresponding Video Segments *Naim, Song, Liu, Huang, Kautz, Luo, and Gildea*	A Neural Network Approach to Context-Sensitive Generation of Conversational Responses *Sordoni, Galley, Auli, Brockett, Ji, Mitchell, Nie, Gao, and Dolan*

Parallel Session 2

Session 2A: Generation and Summarization (Long Papers)

Plaza Ballroom A & B *Chair: Fei Liu*

Transition-Based Syntactic Linearization
Yijia Liu, Yue Zhang, Wanxiang Che, and Bing Qin 14:00–14:25

Syntactic linearization algorithms take a bag of input words and a set of optional constraints, and construct an output sentence and its syntactic derivation simultaneously. The search problem is NP-hard, and the current best results are achieved by bottom-up best-first search.One drawback of the method is low efficiency; and there is no theoretical guarantee that a full sentence can be found within bounded time.We propose an alternative algorithm that constructs output structures from left to right using beam-search.The algorithm is based on incremental parsing algorithms. We extend the transition system so that word ordering is performed in addition to syntactic parsing, resulting in a linearization system that runs in guaranteed quadratic time. In standard evaluations, our system runs an order of magnitude faster than a state-of-the-art baseline using best-first search, with improved accuracies.

Extractive Summarisation Based on Keyword Profile and Language Model
Han Xu, Eric Martin, and Ashesh Mahidadia 14:25–14:50

We present a statistical framework to extract information-rich citation sentences that summarise the main contributions of a scientific paper. In a first stage, we automatically discover salient keywords from a paper's citation summary, keywords that characterise its main contributions. In a second stage, exploiting the results of the first stage, we identify citation sentences that best capture the paper's main contributions. Experimental results show that our approach using methods rooted in quantitative statistics and information theory outperforms the current state-of-the-art systems in scientific paper summarisation.

HEADS: Headline Generation as Sequence Prediction Using an Abstract Feature-Rich Space
Carlos A. Colmenares, Marina Litvak, Amin Mantrach, and Fabrizio Silvestri 14:50–15:15

Automatic headline generation is a sub-task of document summarization with many reported applications. In this study we present a sequence-prediction technique for learning how editors title their news stories. The introduced technique models the problem as a discrete optimization task in a feature-rich space. In this space the global optimum can be found in polynomial time by means of dynamic programming. We train and test our model on an extensive corpus of financial news, and compare it against a number of baselines by using standard metrics from the document summarization domain, as well as some new ones proposed in this work. We also assess the readability and informativeness of the generated titles through human evaluation. The obtained results are very appealing and substantiate the soundness of the approach.

Session 2B: Language and Vision (Long Papers)

Plaza Ballroom D & E *Chair: Yejin Choi*

What's Cookin'? Interpreting Cooking Videos using Text, Speech and Vision
Jonathan Malmaud, Jonathan Huang, Vivek Rathod, Nicholas Johnston, Andrew Rabinovich, and Kevin Murphy 14:00–14:25

We present a novel method for aligning a sequence of instructions to a video of someone carrying out a task. In particular, we focus on the cooking domain, where the instructions correspond to the recipe. Our technique relies on an HMM to align the recipe steps to the (automatically generated) speech transcript. We then refine this alignment using a state-of-the-art visual food detector, based on a deep convolutional neural network. We show that our technique outperforms simpler techniques based on keyword spotting. It also enables interesting applications, such as automatically illustrating recipes with keyframes, and searching within a video for events of interest.

Combining Language and Vision with a Multimodal Skip-gram Model
Angeliki Lazaridou, Nghia The Pham, and Marco Baroni 14:25–14:50

We extend the SKIP-GRAM model of Mikolov et al. (2013a) by taking visual information into account. Like SKIP-GRAM, our multimodal models (MMSKIP-GRAM) build vector-based word representations by learning to predict linguistic contexts in text corpora. However, for a restricted set of words, the models are also exposed to visual representations of the objects they denote (extracted from natural images), and must predict linguistic and visual features jointly. The MMSKIP-GRAM models achieve good performance on a variety of semantic benchmarks. Moreover, since they propagate visual information to all words, we use them to improve image labeling and retrieval in the zero-shot setup, where the test concepts are never seen during model training. Finally, the MMSKIP-GRAM models discover intriguing visual properties of abstract words, paving the way to realistic implementations of embodied theories of meaning.

Discriminative Unsupervised Alignment of Natural Language Instructions with Corresponding Video Segments
Iftekhar Naim, Young C. Song, Qiguang Liu, Liang Huang, Henry Kautz, Jiebo Luo, and Daniel Gildea 14:50–15:15

We address the problem of automatically aligning natural language sentences with corresponding video segments without any direct supervision. Most existing algorithms for integrating language with videos rely on hand-aligned parallel data, where each natural language sentence is manually aligned with its corresponding image or video segment. Recently, fully unsupervised alignment of text with video has been shown to be feasible using hierarchical generative models. In contrast to the previous generative models, we propose three latent-variable discriminative models for the unsupervised alignment task. The proposed discriminative models are capable of incorporating domain knowledge, by adding diverse and overlapping features. The results show that discriminative models outperform the generative models in terms of alignment accuracy.

Session 2C: NLP for Web, Social Media and Social Sciences (Long Papers)

Plaza Ballroom F *Chair: Tong Zhang*

TopicCheck: Interactive Alignment for Assessing Topic Model Stability
Jason Chuang, Margaret E. Roberts, Brandon M. Stewart, Rebecca Weiss, Dustin Tingley, Justin Grimmer, and Jeffrey Heer 14:00–14:25

Content analysis, a widely-applied social science research method, is increasingly being supplemented by topic modeling. However, while the discourse on content analysis centers heavily on reproducibility, computer scientists often focus more on scalability and less on coding reliability, leading to growing skepticism on the usefulness of topic models for automated content analysis. In response, we introduce TopicCheck, an interactive tool for assessing topic model stability. Our contributions are threefold. First, from established guidelines on reproducible content analysis, we distill a set of design requirements on how to computationally assess the stability of an automated coding process. Second, we devise an interactive alignment algorithm for matching latent topics from multiple models, and enable sensitivity evaluation across a large number of models. Finally, we demonstrate that our tool enables social scientists to gain novel insights into three active research questions.

Inferring latent attributes of Twitter users with label regularization
Ehsan Mohammady Ardehaly and Aron Culotta 14:25–14:50

Inferring latent attributes of online users has many applications in public health, politics, and marketing. Most existing approaches rely on supervised learning algorithms, which require manual data annotation and therefore are costly to develop and adapt over time. In this paper, we propose a lightly supervised approach based on label regularization to infer the age, ethnicity, and political orientation of Twitter users. Our approach learns from a heterogeneous collection of soft constraints derived from Census demographics, trends in baby names, and Twitter accounts that are emblematic of class labels. To counteract the imprecision of such constraints, we compare several constraint selection algorithms that optimize classification accuracy on a tuning set. We find that using no user-annotated data, our approach is within 2% of a fully supervised baseline for three of four tasks. Using a small set of labeled data for tuning further improves accuracy on all tasks.

A Neural Network Approach to Context-Sensitive Generation of Conversational Responses
Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan 14:50–15:15

We present a novel response generation system that can be trained end to end on large quantities of unstructured Twitter conversations. A neural network architecture is used to address sparsity issues that arise when integrating contextual information into classic statistical models, allowing the system to take into account previous dialog utterances. Our dynamic-context generative models show consistent gains over both context-sensitive and non-context-sensitive Machine Translation and Information Retrieval baselines.

Session 3 Overview – Monday, June 1, 2015

	Track A	Track B	Track C
	Generation and Summarization (Short Papers) Plaza Ballroom A & B	*Information Extraction and Question Answering (Short Papers)* Plaza Ballroom D & E	*Machine Learning for NLP (Short Papers)* Plaza Ballroom F
3:15	How to Make a Frenemy: Multitape FSTs for Portmanteau Generation *Deri and Knight*	Entity Linking for Spoken Language *Benton and Dredze*	When and why are log-linear models self-normalizing? *Andreas and Klein*
3:30	Aligning Sentences from Standard Wikipedia to Simple Wikipedia *Hwang, Hajishirzi, Ostendorf, and Wu*	Spinning Straw into Gold: Using Free Text to Train Monolingual Alignment Models for Non-factoid Question Answering *Sharp, Jansen, Surdeanu, and Clark*	Deep Multilingual Correlation for Improved Word Embeddings *Lu, Wang, Bansal, Gimpel, and Livescu*
3:45	Inducing Lexical Style Properties for Paraphrase and Genre Differentiation *Pavlick and Nenkova*	Personalized Page Rank for Named Entity Disambiguation *Pershina, He, and Grishman*	Disfluency Detection with a Semi-Markov Model and Prosodic Features *Ferguson, Durrett, and Klein*

Parallel Session 3

Session 3A: Generation and Summarization (Short Papers)

Plaza Ballroom A & B *Chair: Fei Liu*

How to Make a Frenemy: Multitape FSTs for Portmanteau Generation
Aliya Deri and Kevin Knight 15:15–15:30

A portmanteau is a type of compound word that fuses the sounds and meanings of two component words; for example, "frenemy" (friend + enemy) or "smog" (smoke + fog). We develop a system, including a novel multitape FST, that takes an input of two words and outputs possible portmanteaux. Our system is trained on a list of known portmanteaux and their component words, and achieves 45% exact matches in cross-validated experiments.

Aligning Sentences from Standard Wikipedia to Simple Wikipedia
William Hwang, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu 15:30–15:45

This work improves monolingual sentence alignment for text simplification, specifically for text in standard and simple Wikipedia. We introduce a method that improves over past efforts by using a greedy (vs. ordered) search over the document and a word-level se- mantic similarity score based on Wiktionary (vs. WordNet) that also accounts for structural similarity through syntactic dependencies. Experiments show improved performance on a hand-aligned set, with the largest gain coming from structural similarity. Resulting datasets of manually and automatically aligned sentence pairs are made available.

Inducing Lexical Style Properties for Paraphrase and Genre Differentiation
Ellie Pavlick and Ani Nenkova 15:45–16:00

We present an intuitive and effective method for inducing style scores on words and phrases. We exploit signal in a phrase's rate of occurrence across stylistically contrasting corpora, making our method simple to implement and efficient to scale. We show strong results both intrinsically, by correlation with human judgements, and extrinsically, in applications to genre analysis and paraphrasing.

Session 3B: Information Extraction and Question Answering (Short Papers)

Plaza Ballroom D & E *Chair: Yejin Choi*

Entity Linking for Spoken Language
Adrian Benton and Mark Dredze 15:15–15:30

Research on entity linking has considered a broad range of text, including newswire, blogs and web documents in multiple languages. However, the problem of entity linking for spoken language remains unexplored. Spoken language obtained from automatic speech recognition systems poses different types of challenges for entity linking; transcription errors can distort the context, and named entities tend to have high error rates. We propose features to mitigate these errors and evaluate the impact of ASR errors on entity linking using a new corpus of entity linked broadcast news transcripts.

Spinning Straw into Gold: Using Free Text to Train Monolingual Alignment Models for Non-factoid Question Answering
Rebecca Sharp, Peter Jansen, Mihai Surdeanu, and Peter Clark 15:30–15:45

Monolingual alignment models have been shown to boost the performance of question answering systems by "bridging the lexical chasm" between questions and answers. The main limitation of these approaches is that they require semistructured training data in the form of question-answer pairs, which is difficult to obtain in specialized domains or low-resource languages. We propose two inexpensive methods for training alignment models solely using free text, by generating artificial question-answer pairs from discourse structures. Our approach is driven by two representations of discourse: a shallow sequential representation, and a deep one based on Rhetorical Structure Theory. We evaluate the proposed model on two corpora from different genres and domains: one from Yahoo! Answers and one from the biology domain, and two types of non-factoid questions: manner and reason. We show that these alignment models trained directly from discourse structures imposed on free text improve performance considerably over an information retrieval baseline and a neural network language model trained on the same data.

Personalized Page Rank for Named Entity Disambiguation
Maria Pershina, Yifan He, and Ralph Grishman 15:45–16:00

The task of Named Entity Disambiguation is to map entity mentions in the document to their correct entries in some knowledge base. We present a novel graph-based disambiguation approach based on Personalized PageRank (PPR) that combines local and global evidence for disambiguation and effectively filters out noise introduced by in- correct candidates. Experiments show that our method outperforms state-of-the-art approaches by achieving 91.7% in micro- and 89.9% in macroaccuracy on a dataset of 27.8K named entity mentions.

Session 3C: Machine Learning for NLP (Short Papers)

Plaza Ballroom F *Chair: Tong Zhang*

When and why are log-linear models self-normalizing?
Jacob Andreas and Dan Klein 15:15–15:30

Several techniques have recently been pro- posed for training "self-normalized" discriminative models. These attempt to find parameter settings for which unnormalized model scores approximate the true label probability. However, the theoretical properties of such techniques (and of self-normalization generally) have not been investigated. This paper examines the conditions under which we can expect self-normalization to work. We characterize a general class of distributions that admit self-normalization, and prove generalization bounds for procedures that minimize empirical normalizer variance. Motivated by these results, we describe a novel variant of an established procedure for training self-normalized models. The new procedure avoids computing normalizers for most training examples, and decreases training time by as much as factor of ten while preserving model quality.

Deep Multilingual Correlation for Improved Word Embeddings
Ang Lu, Weiran Wang, Mohit Bansal, Kevin Gimpel, and Karen Livescu 15:30–15:45

Word embeddings have been found useful for many NLP tasks, including part-of-speech tagging, named entity recognition, and parsing. Adding multilingual context when learning embeddings can improve their quality, for example via canonical correlation analysis (CCA) on embeddings fromtwo languages. In this paper, we extend this idea to learn deep non-linear transformations of word embeddings of the two languages, using the recently proposed deep canonical correlation analysis. The resulting embeddings, when evaluated on multiple word and bigram similarity tasks, consistently improve over monolingual embeddings and over embeddings transformed with linear CCA.

Disfluency Detection with a Semi-Markov Model and Prosodic Features
James Ferguson, Greg Durrett, and Dan Klein 15:45–16:00

We present a discriminative model for detecting disfluencies in spoken language transcripts. Structurally, our model is a semi-Markov conditional random field with features targeting characteristics unique to speech repairs. This gives a significant performance improvement over standard chain-structured CRFs that have been employed in past work. We then incorporate prosodic features over silences and relative word duration into our semi-CRF model, resulting in further performance gains; moreover, these features are not easily replaced by discrete prosodic indicators such as ToBI breaks. Our final system, the semi-CRF with prosodic information, achieves an F-score of 85.4, which is 1.3 F1 better than the best prior reported F-score on this dataset.

One Minute Madness!

Prior to the poster session, TACL and long-paper poster presenters will be given one minute each to pitch their paper. The poster session will immediately follow these presentations along with a buffet dinner.

Chair: Joel Tetreault

Poster session 1A

Time: *6:00–7:30* Location: Plaza Ballroom A, B, & C

Robust Morphological Tagging with Word Representations
Thomas Müller and Hinrich Schuetze
We present a comparative investigation of word representations for part-of-speech and morphological tagging, focusing on scenarios with considerable differences between training and test data where a robust approach is necessary. Instead of adapting the model towards a specific domain we aim to build a robust model across domains. We developed a test suite for robust tagging consisting of six languages and different domains. We find that representations similar to Brown clusters perform best for POS tagging and that word representations based on linguistic morphological analyzers perform best for morphological tagging.

Multiview LSA: Representation Learning via Generalized CCA
Pushpendre Rastogi, Benjamin Van Durme, and Raman Arora
Multiview LSA (MVLSA) is a generalization of Latent Semantic Analysis (LSA) that supports the fusion of arbitrary views of data and relies on Generalized Canonical Correlation Analysis (GCCA). We present an algorithm for fast approximate computation of GCCA, which when coupled with methods for handling missing values, is general enough to approximate some recent algorithms for inducing vector representations of words. Experiments across a comprehensive collection of test-sets show our approach to be competitive with the state of the art.

Incrementally Tracking Reference in Human/Human Dialogue Using Linguistic and Extra-Linguistic Information
Casey Kennington, Ryu Iida, Takenobu Tokunaga, and David Schlangen
A large part of human communication involves referring to entities in the world and often these entities are objects that are visually present for the interlocutors. A system that aims to resolve such references needs to tackle a complex task: objects and their visual features need to be determined, the referring expressions must be recognised, and extra-linguistic information such as eye gaze or pointing gestures need to be incorporated. Systems that can make use of such information sources exist, but have so far only been tested under very constrained settings, such as WOz interactions. In this paper, we apply to a more complex domain a reference resolution model that works incrementally (i.e., word by word), grounds words with visually present properties of objects (such as shape and size), and can incorporate extra-linguistic information. We find that the model works well compared to previous work on the same data, despite using fewer features. We conclude that the model shows potential for use in a real-time interactive dialogue system.

Digital Leafleting: Extracting Structured Data from Multimedia Online Flyers
Emilia Apostolova, Payam Pourashraf, and Jeffrey Sack
Marketing materials such as flyers and other infographics are a vast online resource. In a number of industries, such as the commercial real estate industry, they are in fact the only authoritative source of information. Companies attempting to organize commercial real estate inventories spend a significant amount of resources on manual data entry of this information. In this work, we propose a method for extracting structured data from free-form commercial real estate flyers in PDF and HTML formats. We modeled the problem as text categorization and Named Entity Recognition (NER) tasks and applied a supervised machine learning approach (Support Vector Machines). Our dataset consists of more than 2,200 commercial real estate flyers and associated manually entered structured data, which was used to automatically create training datasets. Traditionally, text categorization and NER approaches are based on textual information only. However, information in visu-

ally rich formats such as PDF and HTML is often conveyed by a combination of textual and visual features. Large fonts, visually salient colors, and positioning often indicate the most relevant pieces of information. We applied novel features based on visual characteristics in addition to traditional text features and show that performance improved significantly for both the text categorization and NER tasks.

Towards a standard evaluation method for grammatical error detection and correction
Mariano Felice and Ted Briscoe
We present a novel evaluation method for grammatical error correction that addresses problems with previous approaches and scores systems in terms of improvement on the original text. Our method evaluates corrections at the token level using a globally optimal alignment between the source, a system hypothesis and a reference. Unlike the M² Scorer, our method provides scores for both detection and correction and is sensitive to different types of edit operations.

Constraint-Based Models of Lexical Borrowing
Yulia Tsvetkov, Waleed Ammar, and Chris Dyer
Linguistic borrowing is the phenomenon of transferring linguistic constructions (lexical, phonological, morphological, and syntactic) from a "donor" language to a "recipient" language as a result of contacts between communities speaking different languages. Borrowed words are found in all languages, and—in contrast to cognate relationships—borrowing relationships may exist across unrelated languages (for example, about 40% of Swahili's vocabulary is borrowed from Arabic). In this paper, we develop a model of morpho-phonological transformations across languages with features based on universal constraints from Optimality Theory (OT). Compared to several standard—but linguistically naïve—baselines, our OT-inspired model obtains good performance with only a few dozen training examples, making this a cost-effective strategy for sharing lexical information across languages.

Jointly Modeling Inter-Slot Relations by Random Walk on Knowledge Graphs for Unsupervised Spoken Language Understanding
Yun-Nung Chen, William Yang Wang, and Alexander Rudnicky
A key challenge of designing coherent semantic ontology for spoken language understanding is to consider inter-slot relations. In practice, however, it is difficult for domain experts and professional annotators to define a coherent slot set, while considering various lexical, syntactic, and semantic dependencies. In this paper, we exploit the typed syntactic dependency theory for unsupervised induction and filling of semantics slots in spoken dialogue systems. More specifically, we build two knowledge graphs: a slot-based semantic graph, and a word-based lexical graph. To jointly consider word-to-word, word-to-slot, and slot-to-slot relations, we use a random walk inference algorithm to combine the two knowledge graphs, guided by dependency grammars. The experiments show that considering inter-slot relations is crucial for generating a more coherent and compete slot set, resulting in a better spoken language understanding model, while enhancing the interpretability of semantic slots.

Interpreting Compound Noun Phrases Using Web Search Queries
Marius Pasca
A weakly-supervised method is applied to anonymized queries to extract lexical interpretations of compound noun phrases (e.g., "fortune 500 companies"). The interpretations explain the subsuming role ("listed in") that modifiers ("fortune 500") play relative to heads ("companies") within the noun phrases. Experimental results over evaluation sets of noun phrases from multiple sources demonstrate that interpretations extracted from queries have encouraging coverage and precision. The top interpretation extracted is deemed relevant for more than 70% of the noun phrases.

Expanding Paraphrase Lexicons by Exploiting Lexical Variants
Atsushi Fujita and Pierre Isabelle
This study tackles the problem of paraphrase acquisition: achieving high coverage as well as accuracy. Our method first induces paraphrase patterns from given seed paraphrases, exploiting the generality of paraphrases exhibited by pairs of lexical variants, e.g., "amendment" and "amending," in a fully empirical way. It then searches monolingual corpora for new paraphrases that match the patterns. This can extract paraphrases comprising words that are completely different from those of the given seeds. In experiments, our method expanded seed sets by factors of 42 to 206, gaining 84% to 208% more coverage than a previous method that generalizes only identical word forms. Human evaluation through a paraphrase substitution test demonstrated that the newly acquired paraphrases retained reasonable quality, given substantially high-quality seeds.

Lexicon-Free Conversational Speech Recognition with Neural Networks
Andrew Maas, Ziang Xie, Dan Jurafsky, and Andrew Ng

We present an approach to speech recognition that uses only a neural network to map acoustic input to characters, a character-level language model, and a beam search decoding procedure. This approach eliminates much of the complex infrastructure of modern speech recognition systems, making it possible to directly train a speech recognizer using errors generated by spoken language understanding tasks. The system naturally handles out of vocabulary words and spoken word fragments. We demonstrate our approach using the challenging Switchboard telephone conversation transcription task, achieving a word error rate competitive with existing baseline systems. To our knowledge, this is the first entirely neural-network-based system to achieve strong speech transcription results on a conversational speech task. We analyze qualitative differences between transcriptions produced by our lexicon-free approach and transcriptions produced by a standard speech recognition system. Finally, we evaluate the impact of large context neural network character language models as compared to standard n-gram models within our framework.

A Linear-Time Transition System for Crossing Interval Trees
Emily Pitler and Ryan McDonald

We define a restricted class of non-projective trees that 1) covers many natural language sentences; and 2) can be parsed exactly with a generalization of the popular arc-eager system for projective trees (Nivre, 2003). Crucially, this generalization only adds constant overhead in run-time and space keeping the parser's total run-time linear in the worst case. In empirical experiments, our proposed transition-based parser is more accurate on average than both the arc-eager system or the swap-based system, an unconstrained non-projective transition system with a worst-case quadratic runtime (Nivre, 2009).

Data-driven sentence generation with non-isomorphic trees
Miguel Ballesteros, Bernd Bohnet, Simon Mille, and Leo Wanner

Abstract structures from which the generation naturally starts often do not contain any functional nodes, while surface-syntactic structures or a chain of tokens in a linearized tree contain all of them. Therefore, data-driven linguistic generation needs to be able to cope with the projection between non-isomorphic structures that differ in their topology and number of nodes. So far, such a projection has been a challenge in data-driven generation and was largely avoided. We present a fully stochastic generator that is able to cope with projection between non-isomorphic structures. The generator, which starts from PropBank-like structures, consists of a cascade of SVM-classifier based submodules that map in a series of transitions the input structures onto sentences. The generator has been evaluated for English on the Penn-Treebank and for Spanish on the multi-layered Ancora-UPF corpus.

Ontologically Grounded Multi-sense Representation Learning for Semantic Vector Space Models
Sujay Kumar Jauhar, Chris Dyer, and Eduard Hovy

Words are polysemous. However, most approaches to representation learning for lexical semantics assign a single vector to every surface word type. Meanwhile, lexical ontologies such as WordNet provide a source of complementary knowledge to distributional information, including a word sense inventory. In this paper we propose two novel and general approaches for generating sense-specific word embeddings that are grounded in an ontology. The first applies graph smoothing as a post-processing step to tease the vectors of different senses apart, and is applicable to any vector space model. The second adapts predictive maximum likelihood models that learn word embeddings with latent variables representing senses grounded in an specified ontology. Empirical results on lexical semantic tasks show that our approaches effectively captures information from both the ontology and distributional statistics. Moreover, in most cases our sense-specific models outperform other models we compare against.

Subsentential Sentiment on a Shoestring: A Crosslingual Analysis of Compositional Classification
Michael Haas and Yannick Versley

Sentiment analysis has undergone a shift from document-level analysis, where labels express the sentiment of a whole document or whole sentence, to subsentential approaches, which assess the contribution of individual phrases, in particular including the composition of sentiment terms and phrases such as negators and intensifiers. Starting from a small sentiment treebank modeled after the Stanford Sentiment Treebank of Socher et al. (2013), we investigate suitable methods to perform compositional sentiment classification for German in a data-scarce setting, harnessing cross-lingual methods as well as existing general-domain lexical resources.

Using Summarization to Discover Argument Facets in Online Idealogical Dialog
Amita Misra, Pranav Anand, Jean E. Fox Tree, and Marilyn Walker

More and more of the information available on the web is dialogic, and a significant portion of it takes place in online forum conversations about current social and political topics. We aim to develop tools to summarize what these conversations are about. What are the CENTRAL PROPOSITIONS associated with different stances on an issue; what are the abstract objects under discussion that are central to a speaker's argument? How can we recognize that two CENTRAL PROPOSITIONS realize the same FACET of the argument? We hypothesize that the CENTRAL PROPOSITIONS are exactly those arguments that people find most salient, and use human summarization as a probe for discovering them. We describe our corpus of human summaries of opinionated dialogs, then show how we can identify similar repeated arguments, and group them into FACETS across many discussions of a topic. We define a new task, ARGUMENT FACET SIMILARITY (AFS), and show that we can predict AFS with a .54 correlation score, versus an ngram system baseline of .39 and a semantic textual similarity system baseline of .45.

Incorporating Word Correlation Knowledge into Topic Modeling
Pengtao Xie, Diyi Yang, and Eric Xing

This paper studies how to incorporate the external word correlation knowledge to improve the coherence of topic modeling. Existing topic models assume words are generated independently and lack the mechanism to utilize the rich similarity relationships among words to learn coherent topics. To solve this problem, we build a Markov Random Field (MRF) regularized Latent Dirichlet Allocation (LDA) model, which define a MRF on the latent topic layer of LDA to encourage words labeled as similar to share the same topic label. Under our model, the topic assignment of each word is not independent, but rather affected by the topic labels of its correlated words. Similar words have better chance to be put into the same topic due to the regularization of MRF, hence the coherence of topics can be boosted. In addition, our model can accommodate the subtlety that whether two words are similar depends on which topic they appear in, which allows word with multiple senses to be put into different topics properly. We derive a variational inference method to infer the posterior probabilities and learn model parameters and present techniques to deal with the hard-to-compute partition function in MRF. Experiments on two datasets demonstrate the effectiveness of our model.

Active Learning with Rationales for Text Classification
Manali Sharma, Di Zhuang, and Mustafa Bilgic

We present a simple and yet effective approach that can incorporate rationales elicited from annotators into the training of any off-the-shelf classifier. We show that our simple approach is effective for multinomial naive Bayes, logistic regression, and support vector machines. We additionally present an active learning method tailored specifically for the learning with rationales framework.

The Unreasonable Effectiveness of Word Representations for Twitter Named Entity Recognition
Colin Cherry and Hongyu Guo

Named entity recognition (NER) systems trained on newswire perform very badly when tested on Twitter. Signals that were reliable in copy-edited text disappear almost entirely in Twitter's informal chatter, requiring the construction of specialized models. Using well-understood techniques, we set out to improve Twitter NER performance when given a small set of annotated training tweets. To leverage unlabeled tweets, we build Brown clusters and word vectors, enabling generalizations across distributionally similar words. To leverage annotated newswire data, we employ an importance weighting scheme. Taken all together, we establish a new state-of-the-art on two common test sets. Though it is well-known that word representations are useful for NER, supporting experiments have thus far focused on newswire data. We emphasize the effectiveness of representations on Twitter NER, and demonstrate that their inclusion can improve performance by up to 20 F1.

Inferring Temporally-Anchored Spatial Knowledge from Semantic Roles
Eduardo Blanco and Alakananda Vempala

This paper presents a framework to infer spatial knowledge from verbal semantic role representations. First, we generate potential spatial knowledge deterministically. Second, we determine whether it can be inferred and a degree of certainty. Inferences capture that something is located or is not located somewhere, and temporally anchor this information. An annotation effort shows that inferences are ubiquitous and intuitive to humans.

Is Your Anchor Going Up or Down? Fast and Accurate Supervised Topic Models
Thang Nguyen, Jordan Boyd-Graber, Jeffrey Lund, Kevin Seppi, and Eric Ringger

Topic models provide insights into document collections, and their supervised extensions also capture associated document level metadata such as sentiment. However, inferring such models from data is often slow and cannot scale to big data. We build upon the "anchor" method for learning topic models to capture the relationship between metadata and latent topics by extending the vector space representation of word cooccurrence to include metadata specific dimensions. These additional dimensions reveal new anchor words that reflect specific combinations of metadata and topic. We show that these new latent representations predict sentiment as accurately as supervised topic models, and we find these representations more quickly without sacrificing interpretability.

Grounded Semantic Parsing for Complex Knowledge Extraction
Ankur P. Parikh, Hoifung Poon, and Kristina Toutanova

Recently, there has been increasing interest in learning semantic parsers with indirect supervision, but existing work focuses almost exclusively on question answering. Separately, there have been active pursuits in leveraging databases for distant supervision in information extraction, yet such methods are often limited to binary relations and none can handle nested events. In this paper, we generalize distant supervision to complex knowledge extraction, by proposing the first approach to learn a semantic parser for extracting nested event structures without annotated examples, using only a database of such complex events and unannotated text. The key idea is to model the annotations as latent variables, and incorporate a prior that favors semantic parses containing known events. Experiments on the GENIA event extraction dataset show that our approach can learn from and extract complex biological pathway events. Moreover, when supplied with just five example words per event type, it becomes competitive even among supervised systems, outperforming 19 out of 24 teams that participated in the original shared task.

Using External Resources and Joint Learning for Bigram Weighting in ILP-Based Multi-Document Summarization
Chen Li, Yang Liu, and Lin Zhao

Some state-of-the-art summarization systems use integer linear programming (ILP) based methods that aim to maximize the important concepts covered in the summary. These concepts are often obtained by selecting bigrams from the documents. In this paper, we improve such bigram based ILP summarization methods from different aspects. First we use syntactic information to select more important bigrams. Second, to estimate the importance of the bigrams, in addition to the internal features based on the test documents (e.g., document frequency, bigram positions), we propose to extract features by leveraging multiple external resources (such as word embedding from additional corpus,Wikipedia, Dbpedia,Word-Net, SentiWordNet). The bigram weights are then trained discriminatively in a joint learning model that predicts the bigram weights and selects the summary sentences in the ILP framework at the same time. We demonstrate that our system consistently outperforms the prior ILP method on different TAC data sets, and performs competitively compared to other previously reported best results. We also conducted various analysis to show the contribution of different components.

Transforming Dependencies into Phrase Structures
Lingpeng Kong, Alexander M. Rush, and Noah A. Smith

We present a new algorithm for transforming dependency parse trees into phrase-structure parse trees. We cast the problem as struc- tured prediction and learn a statistical model. Our algorithm is faster than traditional phrase- structure parsing and achieves 90.4% English parsing accuracy and 82.4% Chinese parsing accuracy, near to the state of the art on both benchmarks.

Déjà Image-Captions: A Corpus of Expressive Descriptions in Repetition
Jianfu Chen, Polina Kuznetsova, David Warren, and Yejin Choi

We present a new approach to harvesting a large-scale, high quality image-caption corpus that makes a better use of already existing web data with no additional human efforts. The key idea is to focus on Déjà Image-Captions: naturally existing image descriptions that are repeated almost verbatim — by more than one individual for different images. The resulting corpus provides association structure between 4 million images with 180K unique captions, capturing a rich spectrum of everyday narratives including figurative and pragmatic language. Exploring the use of the new corpus, we also present new conceptual tasks of visually situated paraphrasing, creative image captioning, and creative visual paraphrasing.

Improving the Inference of Implicit Discourse Relations via Classifying Explicit Discourse Connectives
Attapol Rutherford and Nianwen Xue

Discourse relation classification is an important component for automatic discourse parsing and natural language understanding. The performance bottleneck of a discourse parser comes from implicit discourse relations, whose discourse connectives are not overtly present. Explicit discourse connectives can potentially be exploited to collect more training data to collect more data and boost the performance. However, using them indiscriminately has been shown to hurt the performance because not all discourse connectives can be dropped arbitrarily. Based on this insight, we investigate the interaction between discourse connectives and the discourse relations and propose the criteria for selecting the discourse connectives that can be dropped independently of the context without changing the interpretation of the discourse. Extra training data collected only by the freely omissible connectives improve the performance of the system without additional features.

Inferring Missing Entity Type Instances for Knowledge Base Completion: New Dataset and Methods
Arvind Neelakantan and Ming-Wei Chang

Most of previous work in knowledge base (KB) completion has focused on the problem of relation extraction. In this work, we focus on the task of inferring missing entity type instances in a KB, a fundamental task for KB competition yet receives little attention. Due to the novelty of this task, we construct a large-scale dataset and design an automatic evaluation methodology. Our knowledge base completion method uses information within the existing KB and external information from Wikipedia. We show that individual methods trained with a global objective that considers unobserved cells from both the entity and the type side gives consistently higher quality predictions compared to baseline methods. We also perform manual evaluation on a small subset of the data to verify the effectiveness of our knowledge base completion methods and the correctness of our proposed automatic evaluation method.

Pragmatic Neural Language Modelling in Machine Translation
Paul Baltescu and Phil Blunsom

This paper presents an in-depth investigation on integrating neural language models in translation systems. Scaling neural language models is a difficult task, but crucial for real-world applications. This paper evaluates the impact on end-to-end MT quality of both new and existing scaling techniques. We show when explicitly normalising neural models is necessary and what optimisation tricks one should use in such scenarios. We also focus on scalable training algorithms and investigate noise contrastive estimation and diagonal contexts as sources for further speed improvements. We explore the trade-offs between neural models and back-off n-gram models and find that neural models make strong candidates for natural language applications in memory constrained environments, yet still lag behind traditional models in raw translation quality. We conclude with a set of recommendations one should follow to build a scalable neural language model for MT.

English orthography is not "close to optimal"
Garrett Nicolai and Grzegorz Kondrak

In spite of the apparent irregularity of the English spelling system, Chomsky and Halle (1968) characterize it as "near optimal". We investigate this assertion using computational techniques and resources. We design an algorithm to generate word spellings that maximize both phonemic transparency and morphological consistency. Experimental results demonstrate that the constructed system is much closer to optimality than the traditional English orthography.

Key Female Characters in Film Have More to Talk About Besides Men: Automating the Bechdel Test
Apoorv Agarwal, Jiehan Zheng, Shruti Kamath, Sriramkumar Balasubramanian, and Shirin Ann Dey

The Bechdel test is a sequence of three questions designed to assess the presence of women in movies. Many believe that because women are seldom represented in film as strong leaders and thinkers, viewers associate weaker stereotypes with women. In this paper, we present a computational approach to automate the task of finding whether a movie passes or fails the Bechdel test. This allows us to study the key differences in language use and in the importance of roles of women in movies that pass the test versus the movies that fail the test. Our experiments confirm that in movies that fail the test, women are in fact portrayed as less-central and less-important characters.

[TACL] Dense Event Ordering with a Multi-Pass Architecture
Nathanael Chambers, Taylor Cassidy, Bill McDowell, and Steven Bethard

The past 10 years of event ordering research has focused on learning partial orderings over document events and time expressions. The most popular corpus, the TimeBank, contains a small subset of the possible ordering graph. Many evaluations follow suit by only testing certain pairs of events (e.g., only main verbs of neighboring sentences). This has led most research to focus on specific learners for partial labelings. This paper attempts to nudge the discussion from identifying some relations to all relations. We present new experiments on strongly connected event graphs that contain ĩ0 times more relations per document than the TimeBank. We also describe a shift away from the single learner to a sieve-based architecture that naturally blends multiple learners into a precision-ranked cascade of sieves. Each sieve adds labels to the event graph one at a time, and earlier sieves inform later ones through transitive closure. This paper thus describes innovations in both approach and task. We experiment on the densest event graphs to date and show a 14% gain over state-of-the-art.

[TACL] Unsupervised Discovery of Biographical Structure from Text
David Bamman and Noah A. Smith

We present a method for discovering abstract event classes in biographies, based on a probabilistic latent-variable model. Taking as input timestamped text, we exploit latent correlations among events to learn a set of event classes (such as "Born", "Graduates high school", and "Becomes citizen"), along with the typical times in a person's life when those events occur. In a quantitative evaluation at the task of predicting a person's age for a given event, we find that our generative model outperforms a strong linear regression baseline, along with simpler variants of the model that ablate some features. The abstract event classes that we learn allow us to perform a large-scale analysis of 242,970 Wikipedia biographies. Though it is known that women are greatly underrepresented on Wikipedia – not only as editors (Wikipedia, 2011) but also as subjects of articles (Reagle and Rhue, 2011) – we find that there is a bias in their characterization as well, with biographies of women containing significantly more emphasis on events of marriage and divorce than biographies of men.

[TACL] Locally Non-Linear Learning for Statistical Machine Translation via Discretization and Structured Regularization
Jonathan H. Clark, Chris Dyer, and Alon Lavie

Linear models, which support efficient learning and inference, are the workhorses of statistical machine translation; however, linear decision rules are less attractive from a modeling perspective. In this work, we introduce a technique for learning arbitrary, rule-local, nonlinear feature transforms that improve model expressivity, but do not sacrifice the efficient inference and learning associated with linear models. To demonstrate the value of our technique, we discard the customary log transform of lexical probabilities and drop the phrasal translation probability in favor of raw counts. We observe that our algorithm learns a variation of a log transform that leads to better translation quality compared to the explicit log transform. We conclude that non-linear responses play an important role in SMT, an observation that we hope will inform the efforts of feature engineers.

[TACL] 2-Slave Dual Decomposition for Generalized High Order CRFs
Xian Qian and Yang Liu

We show that the decoding problem in generalized Higher Order Conditional Random Fields (CRFs) can be decomposed into two parts: one is a tree labeling problem that can be solved in linear time using dynamic programming; the other is a supermodular quadratic pseudo-Boolean maximization problem, which can be solved in cubic time using a minimum cut algorithm. We use dual decomposition to force their agreement. Experimental results on Twitter named entity recognition and sentence dependency tagging tasks show that our method outperforms spanning tree based dual decomposition.

[TACL] SPRITE: Generalizing Topic Models with Structured Priors
Michael J. Paul and Mark Dredze

We introduce SPRITE, a family of topic models that incorporates structure into model priors as a function of underlying components. The structured priors can be constrained to model topic hierarchies, factorizations, correlations, and supervision, allowing SPRITE to be tailored to particular settings. We demonstrate this flexibility by constructing a SPRITE-based model to jointly infer topic hierarchies and author perspective, which we apply to corpora of political debates and online reviews. We show that the model learns intuitive topics, outperforming several other topic models at predictive tasks.

[TACL] Learning Strictly Local Subsequential Functions
Jane Chandlee, Remi Eyraud, and Jeffrey Heinz

We define two proper subclasses of subsequential functions based on the concept of Strict Locality (McNaughton and Papert, 1971; Rogers and Pullum, 2011; Rogers et al., 2013) for formal languages. They are called Input and Output Strictly Local (ISL and OSL). We provide an automata-theoretic characterization of the ISL class and theorems establishing how the classes are related to each other and to Strictly Local languages. We give evidence that local phonological and morphological processes belong to these classes. Finally we provide a learning algorithm which provably identifies the class of ISL functions in the limit from positive data in polynomial time and data. We demonstrate this learning result on appropriately synthesized artificial corpora. We leave a similar learning result for OSL functions for future work and suggest future directions for addressing non-local phonological processes.

[TACL] A sense-topic model for WSI with unsupervised data enrichment
Wang Jing, Mohit Bansal, Kevin Gimpel, Brian D. Ziebart, and Clement T. Yu

Word sense induction (WSI) seeks to automatically discover the senses of a word in a corpus via unsupervised methods. We propose a sense-topic model for WSI, which treats sense and topic as two separate latent variables to be inferred jointly. Topics are informed by the entire document, while senses are informed by the local context surrounding the ambiguous word. We also discuss unsupervised ways of enriching the original corpus in order to improve model performance, including using neural word embeddings and external corpora to expand the context of each data instance. We demonstrate significant improvements over the previous state-of-the-art, achieving the best results reported to date on the SemEval-2013 WSI task.

Poster session 1B

Time: *7:30–9:00* Location: Plaza Ballroom A, B, & C

Modeling Word Meaning in Context with Substitute Vectors
Oren Melamud, Ido Dagan, and Jacob Goldberger

Context representations are a key element in distributional models of word meaning. In contrast to typical representations based on neighboring words, a recently proposed approach suggests to represent a context of a target word by a substitute vector, comprising the potential fillers for the target word slot in that context. In this work we first propose a variant of substitute vectors, which we find particularly suitable for measuring context similarity. Then, we propose a novel model for representing word meaning in context based on this context representation. Our model outperforms state-of-the-art results on lexical substitution tasks in an unsupervised setting.

LCCT: A Semi-supervised Model for Sentiment Classification
Min Yang, Wenting Tu, Ziyu Lu, Wenpeng Yin, and Kam-Pui Chow

Analyzing public opinions towards products, services and social events is an important but challenging task. An accurate sentiment analyzer should take both lexicon-level information and corpus-level information into account. It also needs to exploit the domain-specific knowledge and utilize the common knowledge shared across domains. In addition, we want the algorithm being able to deal with missing labels and learning from incomplete sentiment lexicons. This paper presents a LCCT (Lexicon-based and Corpus-based, Co-Training) model for semi-supervised sentiment classification. The proposed method combines the idea of lexicon-based learning and corpus-based learning in a unified co-training framework. It is capable of incorporating both domain-specific and domain-independent knowledge. Extensive experiments show that it achieves very competitive classification accuracy, even with a small portion of labeled data. Comparing to state-of-the-art sentiment classification methods, the LCCT approach exhibits significantly better performances on a variety of datasets in both English and Chinese.

Empty Category Detection With Joint Context-Label Embeddings
Xun Wang, Katsuhito Sudoh, and Masaaki Nagata

This paper presents a novel technique for empty category (EC) detection using distributed word representations. We formulate it as an annotation task. We explore the hidden layer of a neural network by mapping both the distributed representations of the contexts of ECs and EC types to a low dimensional space. In the testing phase, using the model learned from annotated data, we project the context of a possible EC position to the same space and further compare it with the representations of EC types. The closet EC type is assigned

to the candidate position. Experiments on Chinese Treebank prove the effectiveness of the proposed method. We improve the precision by about 6 point on a subset of Chinese Treebank, which is a new state-of-the-art performance on CTB.

NASARI: a Novel Approach to a Semantically-Aware Representation of Items
José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli

The semantic representation of individual word senses and concepts is of fundamental importance to several applications in Natural Language Processing. To date, concept modeling techniques have in the main based their representation either on lexicographic resources, such as WordNet, or on encyclopedic resources, such as Wikipedia. We propose a vector representation technique that combines the complementary knowledge of both these types of resource. Thanks to its use of explicit semantics combined with a novel cluster-based dimensionality reduction and an effective weighting scheme, our representation attains state-of-the-art performance on multiple datasets in two standard benchmarks: word similarity and sense clustering. We are releasing our vector representations at http://lcl.uniroma1.it/nasari/.

Multi-Target Machine Translation with Multi-Synchronous Context-free Grammars
Graham Neubig, Philip Arthur, and Kevin Duh

We propose a method for simultaneously translating from a single source language to multiple target languages T1, T2, etc. The motivation behind this method is that if we only have a weak language model for T1 and translations in T1 and T2 are associated, we can use the information from a strong language model over T2 to disambiguate the translations in T1, providing better translation results. As a specific framework to realize multi-target translation, we expand the formalism of synchronous context-free grammars to handle multiple targets, and describe methods for rule extraction, scoring, pruning, and search with these models. Experiments find that multi-target translation with a strong language model in a similar second target language can provide gains of up to 0.8-1.5 BLEU points.

Using Zero-Resource Spoken Term Discovery for Ranked Retrieval
Jerome White, Douglas Oard, Aren Jansen, Jiaul Paik, and Rashmi Sankepally

Research on ranked retrieval of spoken content has assumed the existence of some automated (word or phonetic) transcription. Recently, however, methods have been demonstrated for matching spoken terms to spoken content without the need for language-tuned transcription. This paper describes the first application of such techniques to ranked retrieval, evaluated using a newly created test collection. Both the queries and the collection to be searched are based on Gujarati produced naturally by native speakers; relevance assessment was performed by other native speakers of Gujarati. Ranked retrieval is based on fast acoustic matching that identifies a deeply nested set of matching speech regions, coupled with ways of combining evidence from those matching regions. Results indicate that the resulting ranked lists may be useful for some practical similarity-based ranking tasks.

Sign constraints on feature weights improve a joint model of word segmentation and phonology
Mark Johnson, Joe Pater, Robert Staubs, and Emmanuel Dupoux

This paper describes a joint model of word segmentation and phonological alternations, which takes unsegmented utterances as input and infers word segmentations and underlying phonological representations. The model is a Maximum Entropy or log-linear model, which can express a probabilistic version of Optimality Theory (OT; Prince 2004), a standard phonological framework. The features in our model are inspired by OT's Markedness and Faithfulness constraints. Following the OT principle that such features indicate "violations", we require their weights to be non-positive. We apply our model to a modified version of the Buckeye corpus (Pitt 2007) in which the only phonological alternations are deletions of word-final /d/ and /t/ segments. The model sets a new state-of-the-art for this corpus for word segmentation, identification of underlying forms, and identification of /d/ and /t/ deletions. We also show that the OT-inspired sign constraints on feature weights are crucial for accurate identification of deleted /d/s; without them our model posits approximately 10 times more deleted underlying /d/s than appear in the manually annotated data.

Semi-Supervised Word Sense Disambiguation Using Word Embeddings in General and Specific Domains
Kaveh Taghipour and Hwee Tou Ng

One of the weaknesses of current supervised word sense disambiguation (WSD) systems is that they only treat a word as a discrete entity. However, a continuous-space representation of words (word embeddings) can provide valuable information and thus improve generalization accuracy. Since word embeddings are typically obtained from unlabeled data using unsupervised methods, this method can be seen as a semi-supervised word sense disambiguation approach. This paper investigates two ways of incorporating word embeddings in a word sense disambiguation setting and evaluates these two methods on some SensEval/SemEval lexical sample and all-words tasks and also a domain-specific lexical sample task. The obtained results show that such representations consistently improve the accuracy of the selected supervised WSD system. Moreover, our experiments on a domain-specific dataset show that our supervised baseline system beats the best knowledge-based systems by a large margin.

Model Invertibility Regularization: Sequence Alignment With or Without Parallel Data
Tomer Levinboim, Ashish Vaswani, and David Chiang

We present Model Invertibility Regularization MIR, a method that jointly trains two directional sequence alignment models, one in each direction, and takes into account the invertibility of the alignment task. By coupling the two models through their parameters (as opposed to through their inferences, as in Liang et al.'s Alignment by Agreement (ABA), and Ganchev et al.'s Posterior Regularization (PostCAT)), our method seamlessly extends to all IBM-style word alignment models as well as to alignment without parallel data. Our proposed algorithm is mathematically sound and inherits convergence guarantees from EM. We evaluate MIR on two tasks: (1) On word alignment, applying MIR on fertility based models we attain higher F-scores than ABA and PostCAT. (2) On Japanese-to-English back-transliteration without parallel data, applied to the decipherment model of Ravi and Knight, MIR learns sparser models that close the gap in whole-name error rate by 33% relative to a model trained on parallel data, and further, beats a previous approach by Mylonakis et al.

Continuous Space Representations of Linguistic Typology and their Application to Phylogenetic Inference
Yugo Murawaki

For phylogenetic inference, linguistic typology is a promising alternative to lexical evidence because it allows us to compare an arbitrary pair of languages. A challenging problem with typology-based phylogenetic inference is that the changes of typological features over time are less intuitive than those of lexical features. In this paper, we work on reconstructing typologically natural ancestors To do this, we leverage dependencies among typological features. We first represent each language by continuous latent components that capture feature dependencies. We then combine them with a typology evaluator that distinguishes typologically natural languages from other possible combinations of features. We perform phylogenetic inference in the continuous space and use the evaluator to ensure the typological naturalness of inferred ancestors. We show that the proposed method reconstructs known language families more accurately than baseline methods. Lastly, assuming the monogenesis hypothesis, we attempt to reconstruct a common ancestor of the world's languages.

Diamonds in the Rough: Event Extraction from Imperfect Microblog Data
Ander Intxaurrondo, Eneko Agirre, Oier Lopez de Lacalle, and Mihai Surdeanu

We introduce a distantly supervised event ex- traction approach that extracts complex event templates from microblogs. We show that this near real-time data source is more challeng- ing than news because it contains information that is both approximate (e.g., with values that are close but different from the gold truth) and ambiguous (due to the brevity of the texts), impacting both the evaluation and extraction methods. For the former, we propose a novel, "soft", F1 metric that incorporates similarity between extracted fillers and the gold truth, giving partial credit to different but similar values. With respect to extraction method- ology, we propose two extensions to the dis- tant supervision paradigm: to address approx- imate information, we allow positive training examples to be generated from information that is similar but not identical to gold values; to address ambiguity, we aggregate contexts across tweets discussing the same event. We evaluate our contributions on the complex do- main of earthquakes, with events with up to 20 arguments. Our results indicate that, despite their simplicity, our contributions yield a statistically-significant improvement of 25% (relative) over a strong distantly-supervised system. The dataset containing the knowledge base, relevant tweets and manual annotations is publicly available.

I Can Has Cheezburger? A Nonparanormal Approach to Combining Textual and Visual Information for Predicting and Generating Popular Meme Descriptions
William Yang Wang and Miaomiao Wen

The advent of social media has brought Internet memes, a unique social phenomenon, to the front stage of the Web. Embodied in the form of images with text descriptions, little do we know about the "language of memes". In this paper, we statistically study the correlations among popular memes and their wordings, and generate meme descriptions from raw images. To do this, we take a multimodal approach—we propose a robust nonparanormal model to learn the stochastic dependencies among the image, the candidate descriptions, and the popular votes. In experiments, we show that combining text and vision helps identifying popular meme descriptions; that our nonparanormal model is able to learn dense and continuous vision features jointly with sparse and discrete text features in a principled manner, outperforming various competitive baselines; that our system can generate meme descriptions using a simple pipeline.

Unsupervised Dependency Parsing: Let's Use Supervised Parsers
Phong Le and Willem Zuidema

We present a self-training approach to unsupervised dependency parsing that reuses existing supervised and unsupervised parsing algorithms. Our approach, called 'iterated reranking' (IR), starts with dependency trees generated by an unsupervised parser, and iteratively improves these trees using the richer probability models used in supervised parsing that are in turn trained on these trees. Our system achieves 1.8% accuracy higher than the state-of-the-part parser of Spitkovsky et al. (2013) on the WSJ corpus.

A Transition-based Algorithm for AMR Parsing
Chuan Wang, Nianwen Xue, and Sameer Pradhan

We present a two-stage framework to parse a sentence into its Abstract Meaning Representation (AMR). We first use a dependency parser to generate a dependency tree for the sentence. In the second stage, we design a novel transition-based algorithm that transforms the dependency tree to an AMR graph. There are several advantages with this approach. First, the dependency parser can be trained on a training set much larger than the training set for the tree-to-graph algorithm, resulting in a more accurate AMR parser overall. Our parser yields an improvement of 5% absolute in F-measure over the best previous result. Second, the actions that we design are linguistically intuitive and capture the regularities in the mapping between the dependency structure and the AMR of a sentence. Third, our parser runs in nearly linear time in practice in spite of a worst-case complexity of O(n²).

The Geometry of Statistical Machine Translation
Aurelien Waite and Bill Byrne

Most modern statistical machine translation systems are based on linear statistical models. One extremely effective method for estimating the model parameters is minimum error rate training (MERT), which is an efficient form of line optimisation adapted to the highly non-linear objective functions used in machine translation. We describe a polynomial-time generalisation of line optimisation that computes the error surface over a plane embedded in parameter space. The description of this algorithm relies on convex geometry, which is the mathematics of polytopes and their faces. Using this geometric representation of MERT we investigate whether the optimisation of linear models is tractable in general. Previous work on finding optimal solutions in MERT (Galley and Quirk, 2011) established a worst-case complexity that was exponential in the number of sentences, in contrast we show that exponential dependence in the worst-case complexity is mainly in the number of features. Although our work is framed with respect to MERT, the convex geometric description is also applicable to other error-based training methods for linear models. We believe our analysis has important ramifications because it suggests that the current trend in building statistical machine translation systems by introducing a very large number of sparse features is inherently not robust.

Unsupervised Multi-Domain Adaptation with Feature Embeddings
Yi Yang and Jacob Eisenstein

Representation learning is the dominant technique for unsupervised domain adaptation, but existing approaches have two major weaknesses. First, they often require the specification of "pivot features" that generalize across domains, which are selected by task-specific heuristics. We show that a novel but simple feature embedding approach provides better performance, by exploiting the feature template structure common in NLP problems. Second, unsupervised domain adaptation is typically treated as a task of moving from a single source to a single target domain. In reality, test data may be diverse, relating to the training data in some ways but not others. We propose an alternative formulation, in which each instance has a vector of domain attributes, can be used to learn distill the domain-invariant properties of each feature.

Latent Domain Word Alignment for Heterogeneous Corpora
Hoang Cuong and Khalil Sima'an

This work focuses on the insensitivity of existing word alignment models to domain differences, which often yields suboptimal results on large heterogeneous data. A novel latent domain word alignment model is proposed, which induces domain-conditioned lexical and alignment statistics. We propose to train the model on a heterogeneous corpus under partial supervision, using a small number of seed samples from different domains. The seed samples allow estimating sharper, domain-conditioned word alignment statistics for sentence pairs. Our experiments show that the derived domain-conditioned statistics, once combined together, produce notable improvements both in word alignment accuracy and in translation accuracy of their resulting SMT systems.

Extracting Human Temporal Orientation from Facebook Language
H. Andrew Schwartz, Gregory Park, Maarten Sap, Evan Weingarten, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Jonah Berger, Martin Seligman, and Lyle Ungar

People vary widely in their temporal orientation — how often they emphasize the past, present, and future — and this affects their finances, health, and happiness. Traditionally, temporal orientation has been assessed by self-report questionnaires. In this paper, we develop a novel behavior-based assessment using human language on Facebook. We first create a past, present, and future message classifier, engineering features and evaluating a variety of classification techniques. Our message classifier achieves an accuracy of 71.8%, compared with 52.8% from the most frequent class and 58.6% from a model based entirely on time expression features. We quantify a users' overall temporal orientation based on their distribution of messages and validate it against known human correlates: conscientiousness, age, and gender. We then explore social scientific questions, finding novel associations with the factors openness to experience, satisfaction with life, depression, IQ, and one's number of friends. Further, demonstrating how one can track orientation over time, we find differences in future orientation around birthdays.

Cost Optimization in Crowdsourcing Translation: Low cost translations made even cheaper
Mingkun Gao, Wei Xu, and Chris Callison-Burch

Crowdsourcing makes it possible to create translations at much lower cost than hiring professional translators. However, it is still expensive to obtain the millions of translations that are needed to train statistical machine translation systems. We propose two mechanisms to reduce the cost of crowdsourcing while maintaining high translation quality. First, we develop a method to reduce redundant translations. We train a linear model to evaluate the translation quality on a sentence-by-sentence basis, and fit a threshold between acceptable and unacceptable translations. Unlike past work, which always paid for a fixed number of translations for each source sentence and then chose the best from them, we can stop earlier and pay less when we receive a translation that is good enough. Second, we introduce a method to reduce the pool of translators by quickly identifying bad translators after they have translated only a few sentences. This also allows us to rank translators, so that we re-hire only good translators to reduce cost.

An In-depth Analysis of the Effect of Text Normalization in Social Media
Tyler Baldwin and Yunyao Li

Recent years have seen increased interest in text normalization in social media, as the informal writing styles found in Twitter and other social media data often cause problems for NLP applications. Unfortunately, most current approaches narrowly regard the normalization task as a "one size fits all" task of replacing non-standard words with their standard counterparts. In this work we build a taxonomy of normalization edits and present a study of normalization to examine its effect on three different downstream applications (dependency parsing, named entity recognition, and text-to-speech synthesis). The results suggest that how the normalization task should be viewed is highly dependent on the targeted application. The results also show that normalization must be thought of as more than word replacement in order to produce results comparable to those seen on clean text.

Multitask Learning for Adaptive Quality Estimation of Automatically Transcribed Utterances
José G. C. de Souza, Hamed Zamani, Matteo Negri, Marco Turchi, and Falavigna Daniele

We investigate the problem of predicting the quality of automatic speech recognition (ASR) output under the following rigid constraints: i) reference transcriptions are not available, ii) confidence information about the system that produced the transcriptions is not accessible, and iii) training and test data come from multiple domains. To cope with these constraints (typical of the constantly increasing amount of automatic transcrip-

tions that can be found on the Web), we propose a domain-adaptive approach based on multitask learning. Different algorithms and strategies are evaluated with English data coming from four domains, showing that the proposed approach can cope with the limitations of previously proposed single task learning methods.

A Dynamic Programming Algorithm for Tree Trimming-based Text Summarization
Masaaki Nishino, Norihito Yasuda, Tsutomu Hirao, Shin-ichi Minato, and Masaaki Nagata

Tree trimming is the problem of extracting an optimal subtree from an input tree, and sentence extraction and sentence compression methods can be formulated and solved as tree trimming problems. Previous approaches require integer linear programming (ILP) solvers to obtain exact solutions. The problem of this approach is that ILP solvers are black-boxes and have no theoretical guarantee as to their computation complexity. We propose a dynamic programming (DP) algorithm for tree trimming problems whose running time is O(NL log N), where N is the number of tree nodes and L is the length limit. Our algorithm exploits the zero-suppressed binary decision diagram (ZDD), a data structure that represents a family of sets as a directed acyclic graph, to represent the set of subtrees in a compact form; the structure of ZDD permits the application of DP to obtain exact solutions, and our algorithm is applicable to different tree trimming problems. Moreover, experiments show that our algorithm is faster than state-of-the-art ILP solvers, and that it scales well to handle large summarization problems.

Sentiment after Translation: A Case-Study on Arabic Social Media Posts
Mohammad Salameh, Saif Mohammad, and Svetlana Kiritchenko

When text is translated from one language into another, sentiment is preserved to varying degrees. In this paper, we use Arabic social media posts as stand-in for source language text, and determine loss in sentiment predictability when they are translated into English, manually and automatically. As benchmarks, we use manually and automatically determined sentiment labels of the Arabic texts. We show that sentiment analysis of English translations of Arabic texts produces competitive results, w.r.t. Arabic sentiment analysis. We discover that even though translation significantly reduces the human ability to recover sentiment, automatic sentiment systems are still able to capture sentiment information from the translations.

Corpus-based discovery of semantic intensity scales
Chaitanya Shivade, Marie-Catherine de Marneffe, Eric Fosler-Lussier, and Albert M. Lai

Gradable terms such as brief, lengthy and extended illustrate varying degrees of a scale and can therefore participate in comparative constructs. Knowing the set of words that can be compared on the same scale and the associated ordering between them (brief < lengthy < extended) is very useful for a variety of lexical semantic tasks. Current techniques to derive such an ordering rely on WordNet to determine which words belong on the same scale and are limited to adjectives. Here we describe an extension to recent work: we investigate a fully automated pipeline to extract gradable terms from a corpus, group them into clusters reflecting the same scale and establish an ordering among them. This methodology reduces the amount of required handcrafted knowledge, and can infer gradability of words independent of their part of speech. Our approach infers an ordering for adjectives with comparable performance to previous work, but also for adverbs with an accuracy of 71%. We find that the technique is useful for inferring such rankings among words across different domains, and present an example using biomedical text.

Dialogue focus tracking for zero pronoun resolution
Sudha Rao, Allyson Ettinger, Hal Daumé III, and Philip Resnik

We take a novel approach to zero pronoun resolution in Chinese: our model explicitly tracks the flow of focus in a discourse. Our approach, which generalizes to deictic references, is not reliant on the presence of overt noun phrase antecedents to resolve to, and allows us to address the large percentage of "non-anaphoric" pronouns filtered out in other approaches. We furthermore train our model using readily available parallel Chinese/English corpora, allowing for training without hand-annotated data. Our results demonstrate improvements on two test sets, as well as the usefulness of linguistically motivated features.

Solving Hard Coreference Problems
Haoruo Peng, Daniel Khashabi, and Dan Roth

Coreference resolution is a key problem in natural language understanding that still escapes reliable solutions. One fundamental difficulty has been that of resolving instances involving pronouns since they often require deep language understanding and use of background knowledge. In this paper we propose an algorithmic solution that involves a new representation for the knowledge required to address hard coreference problems,

along with a constrained optimization framework that uses this knowledge in coreference decision making. Our representation, Predicate Schemas, is instantiated with knowledge acquired in an unsupervised way, and is compiled automatically into constraints that impact the coreference decision. We present a general coreference resolution system that significantly improves state-of-the-art performance on hard, Winograd-style, pronoun resolution cases, while still performing at the state-of-the-art level on standard coreference resolution datasets.

[TACL] Reasoning about Quantities in Natural Language
Subhro Roy, Tim Vieira, and Dan Roth

Little work from the Natural Language Processing community has targeted the role of quantities in Natural Language Understanding. This paper takes some key steps towards facilitating reasoning about quantities expressed in natural language. We investigate two different tasks of numerical reasoning. First, we consider Quantity Entailment, a new task formulated to understand the role of quantities in general textual inference tasks. Second, we consider the problem of automatically understanding and solving elementary school math word problems. In order to address these quantitative reasoning problems we first develop a computational approach which we show to successfully recognize and normalize textual expressions of quantities. We then use these capabilities to further develop algorithms to assist reasoning in the context of the aforementioned tasks.

[TACL] Learning Constraints for Information Structure Analysis of Scientific Documents
Yufan Guo, Roi Reichart, and Anna Korhonen

Inferring the information structure of scientific documents is useful for many NLP applications. Existing approaches to this task require substantial human effort. We propose a framework for constraint learning that reduces human involvement considerably. Our model uses topic models to identify latent topics and their key linguistic features in input documents, induces constraints from this information and maps sentences to their dominant information structure categories through a constrained unsupervised model. When the induced constraints are combined with a fully unsupervised model, the resulting model challenges existing lightly supervised feature-based models as well as unsupervised models that use manually constructed declarative knowledge. Our results demonstrate that useful declarative knowledge can be learned from data with very limited human involvement.

SRW Poster Session

Time: *6:00–9:00* Location: Plaza Ballroom A, B, & C

Cache-Augmented Latent Topic Language Models for Speech Retrieval
Jonathan Wintrode

We aim to improve speech retrieval performance by augmenting traditional N-gram language models with different types of topic context. We present a latent topic model framework that treats documents as arising from an underlying topic sequence combined with a cache-based repetition model. We analyze our proposed model both for its ability to capture word repetition via the cache and for its suitability as a language model for speech recognition and retrieval. We show this model, augmented with the cache, captures intuitive repetition behavior across languages and exhibits lower perplexity than regular LDA on held out data in multiple languages. Lastly, we show that our joint model improves speech retrieval performance beyond N-grams or latent topics alone, when applied to a term detection task in all languages considered.

Reliable Lexical Simplification for Non-Native Speakers
Gustavo Paetzold

Lexical Simplification is the task of modifying the lexical content of complex sentences in order to make them simpler. Due to the lack of reliable resources available for the task, most existing approaches have difficulties producing simplifications which are grammatical and that preserve the meaning of the original text. In order to improve on the state-of-the-art of this task, we propose user studies with non-native speakers, which will result in new, sizeable datasets, as well as novel ways of performing Lexical Simplification. The results of our first experiments show that new types of classifiers, along with the use of additional resources such as spoken text language models, produce the state-of-the-art results for the Lexical Simplification task of SemEval-2012.

Analyzing Newspaper Crime Reports for Identification of Safe Transit Paths
Vasu Sharma, Rajat Kulshreshtha, Puneet Singh, Nishant Agrawal, and Akshay Kumar

In the present situation, where every day we come across one or the other crime, it is of prime importance to ensure ones safety. Any measure which could ensure ones safety is certain to pay for itself. An automated system which can suggest the safest path between any two locations in the city avoiding crime prone areas will be of utmost importance for the society. In this paper, we propose a method to find the safest path between two locations, based on the geographical model of crime intensities. We consider the police records and news articles for finding crime density of different areas of the city. It is essential to consider news articles as there is a significant delay in updating police crime records. We address this problem by updating the crime intensities based on current news feeds. Based on the updated crime intensities, we identify the safest path. It is this real time updation of crime intensities which makes our model way better than the models that are presently in use. Our model would also inform the user of crime sprees and spurt of crimes in a particular area thereby ensuring that user avoids these crime hot spots.

Relation extraction pattern ranking using word similarity
Konstantinos Lambrou-Latreille

Our thesis proposal aims at integrating word similarity measures in pattern ranking for relation extraction bootstrapping algorithms. We note that although many contributions have been done on pattern ranking schemas, few explored the use of word-level semantic similarity. Our hypothesis is that word similarity would allow better pattern comparison and better pattern ranking, resulting in less semantic drift commonly problematic in bootstrapping algorithms. In this paper, as a first step into this research, we explore different pattern representations, various existing pattern ranking approaches and some word similarity measures. We also present a methodology and evaluation approach to test our hypothesis.

Towards a Better Semantic Role Labeling of Complex Predicates
Glorianna Jagfeld and Lonneke van der Plas

We propose a way to automatically improve the annotation of verbal complex predicates in PropBank which until now has been treating language mostly in a compositional manner. In order to minimize the manual re-annotation effort, we build on the recently introduced concept of aliasing complex predicates to existing PropBank rolesets which encompass the same meaning and argument structure. We suggest to find aliases automatically by applying a multilingual distributional model that uses the translations of simple and complex predicates as features. Furthermore, we set up an annotation effort to obtain a frequency balanced, realistic test set for this task. Our method reaches an accuracy of 44% on this test set and 72% for the more frequent test items in a lenient evaluation, which is not far from the upper bounds from human annotation.

Exploring Relational Features and Learning under Distant Supervision for Information Extraction Tasks
Ajay Nagesh

Information Extraction (IE) has become an indispensable tool in our quest to handle the data deluge of the information age. IE can broadly be classified into Named-entity Recognition (NER) and Relation Extraction (RE). In this thesis, we view the task of IE as finding patterns in unstructured data, which can either take the form of features and/or be specified by constraints. In NER, we study the categorization of complex relational features and outline methods to learn feature combinations through induction. We demonstrate the efficacy of induction techniques in learning : i) rules for the identification of named entities in text — the novelty is the application of induction techniques to learn in a very expressive declarative rule language ii) a richer sequence labeling model — enabling optimal learning of discriminative features. In RE, our investigations are in the paradigm of distant supervision, which facilitates the creation of large albeit noisy training data. We devise an inference framework in which constraints can be easily specified in learning relation extractors. In addition, we reformulate the learning objective in a max-margin framework. To the best of our knowledge, our formulation is the first to optimize multi-variate non-linear performance measures such as F-β for a latent variable structure prediction task.

Entity/Event-Level Sentiment Detection and Inference
Lingjia Deng

Most of the work in sentiment analysis and opinion mining focuses on extracting explicit sentiments. Opinions may be expressed implicitly via inference rules over explicit sentiments. In this thesis, we incorporate the inference rules as constraints in joint prediction models, to develop an entity/event-level sentiment analysis system which aims at detecting both explicit and implicit sentiments expressed among entities and events in

the text, especially focusing on but not limited to sentiments toward events that positively or negatively affect entities (+/-effect events).

Initial Steps for Building a Lexicon of Adjectives with Scalemates
Bryan Wilkinson

This paper describes work in progress to use clustering to create a lexicon of words that engage in the lexico-semantic relationship known as grading. While other resources like thesauri and taxonomies exist detailing relationships such as synonymy, antonymy, and hyponymy, we do not know of any thorough resource for grading. This work focuses on identifying the words that may participate in this relationship, paving the way for the creation of a true grading lexicon later.

A Preliminary Evaluation of the Impact of Syntactic Structure in Semantic Textual Similarity and Semantic Relatedness Tasks
Ngoc Phuoc An Vo and Octavian Popescu

The well related tasks of evaluating the Semantic Textual Similarity and Semantic Relatedness have been under a special attention in NLP community. Many different approaches have been proposed, implemented and evaluated at different levels, such as lexical similarity, word/string/POS tags overlapping, semantic modeling (LSA, LDA), etc. However, at the level of syntactic structure, it is not clear how significant it contributes to the overall accuracy. In this paper, we make a preliminary evaluation of the impact of the syntactic structure in the tasks by running and analyzing the results from several experiments regarding to how syntactic structure contributes to solving the tasks.

Benchmarking Machine Translated Sentiment Analysis for Arabic Tweets
Eshrag Refaee and Verena Rieser

Traditional approaches to Sentiment Analysis (SA) rely on large annotated data sets or wide-coverage sentiment lexica, and as such often perform poorly on under-resourced languages. This paper presents empirical evidence of an efficient SA approach using freely available machine translation (MT) systems to translate Arabic tweets to English, which we then label for sentiment using a state-of-the-art English SA system. We show that this approach significantly outperforms a number of standard approaches on a gold-standard heldout data set, and performs equally well compared to more cost-intense methods with 76% accuracy. This confirms MT-based SA as a cheap and effective alternative to building a fully fledged SA system when dealing with under-resourced languages.

Learning Kernels for Semantic Clustering: A Deep Approach
Ignacio Arroyo-Fernández

In this thesis proposal we present a novel semantic embedding method, which aims at consistently performing semantic clustering at sentence level. Taking into account special aspects of Vector Space Models (VSMs), we propose to learn reproducing kernels in classification tasks. By this way, capturing spectral features from data is possible. These features make it theoretically plausible to model semantic similarity criteria in Hilbert spaces, i.e. the embedding spaces. We could improve the semantic assessment over embeddings, which are criterion-derived representations from traditional semantic vectors. The learned kernel could be easily transferred to clustering methods, where the Multi-Class Imbalance Problem is considered (e.g. semantic clustering of definitions of terms).

Narrowing the Loop: Integration of Resources and Linguistic Dataset Development with Interactive Machine Learning
Seid Muhie Yimam

This thesis proposal sheds light on the role of interactive machine learning and implicit user feedback for manual annotation tasks and semantic writing aid applications. First we focus on the cost-effective annotation of training data using an interactive machine learning approach by conducting an experiment for sequence tagging of German named entity recognition. To show the effectiveness of the approach, we further carry out a sequence tagging task on Amharic part-of-speech and are able to significantly reduce time used for annotation. The second research direction is to systematically integrate different NLP resources for our new semantic writing aid tool using again an interactive machine learning approach to provide contextual paraphrase suggestions. We develop a baseline system where three lexical resources are combined to provide paraphrasing in context and show that combining resources is a promising direction.

Relation Extraction from Community Generated Question-Answer Pairs
Denis Savenkov, Wei-Lwun Lu, Jeff Dalton, and Eugene Agichtein

Community question answering (CQA) websites contain millions of question and answer (QnA) pairs that represent real users' interests. Traditional methods for relation extraction from natural language text operate over individual sentences. However answer text is sometimes hard to understand without knowing the question, e.g., it may not name the subject or relation of the question. This work presents a novel model for relation extraction from CQA data, which uses discourse of QnA pairs to predict relations between entities mentioned in question and answer sentences. Experiments on 2 publicly available datasets demonstrate that the model can extract from ~20% to ~40% additional relation triples, not extracted by existing sentence-based models.

Detecting Translation Direction: A Cross-Domain Study
Sauleh Eetemadi and Kristina Toutanova

Parallel corpora are constructed by taking a document authored in one language and translating it into another language. However, the information about the authored and translated sides of the corpus is usually not preserved. When available, this information can be used to improve statistical machine translation. Existing statistical methods for translation direction detection have low accuracy when applied to the realistic out-of-domain setting, especially when the input texts are short. Our contributions in this work are three fold: 1) We develop a multi-corpus parallel dataset with translation direction labels at the sentence level, 2) we perform a comparative evaluation of previously introduced features for translation direction detection in a cross-domain setting and 3) we generalize a previously introduced type of features to outperform the best previously proposed features in detecting translation direction and achieve 0.80 precision with 0.85 recall.

Improving the Translation of Discourse Markers for Chinese into English
David Steele

Discourse markers (DMs) are ubiquitous cohesive devices used to connect what is said or written. However, across languages there is divergence in their usage, placement, and frequency, which is considered to be a major problem for machine translation (MT). This paper presents an overview of a proposed thesis, exploring the difficulties around DMs in MT, with a focus on Chinese and English. The thesis will examine two main areas: modelling cohesive devices within sentences and modelling discourse relations (DRs) across sentences. Initial experiments have shown promising results for building a prediction model that uses linguistically inspired features to help improve word alignments with respect to the implicit use of cohesive devices, which in turn leads to improved hierarchical phrase based MT.

Discourse and Document-level Information for Evaluating Language Output Tasks
Carolina Scarton

Evaluating the quality of language output tasks such as Machine Translation (MT) and Automatic Summarisation (AS) is a challenging topic in Natural Language Processing (NLP). Recently, techniques focusing only on the use of outputs of the systems and source information have been investigated. In MT, this is referred to as Quality Estimation (QE), an approach based on using machine learning techniques to predict the quality of unseen data, generalising from a few labelled data points. Traditional QE research addresses sentence-level QE evaluation and prediction, disregarding document-level information. Document-level QE requires a different set up from sentence-level, which makes the study of appropriate quality scores, features and models necessary. Our aim is to explore document-level QE of MT, focusing on discourse information. However, the findings of this research can improve other NLP tasks, such as AS.

Speeding Document Annotation with Topic Models
Forough Poursabzi-Sangdeh and Jordan Boyd-Graber

Document classification and topic models are useful tools for managing and understanding large corpora. Topic models are used to uncover underlying semantic and structure of document collections. Categorizing large collection of documents requires hand-labeled training data, which is time consuming and needs human expertise. We believe engaging user in the process of document labeling helps reduce annotation time and address user needs. We present an interactive tool for document labeling. We use topic models to help users in this procedure. Our preliminary results show that users can more effectively and efficiently apply labels to documents using topic model information.

Lifelong Machine Learning for Topic Modeling and Beyond
Zhiyuan Chen

Machine learning has been popularly used in numerous natural language processing tasks. However, most ma-

chine learning models are built using a single dataset. This is often referred to as one-shot learning. Although this one-shot learning paradigm is very useful, it will never make an NLP system understand the natural language because it does not accumulate knowledge learned in the past and make use of the knowledge in future learning and problem solving. In this thesis proposal, I first present a survey of lifelong machine learning (LML). I then narrow down to one specific NLP task, i.e., topic modeling. I propose several approaches to apply lifelong learning idea in topic modeling. Such capability is essential to make an NLP system versatile and holistic.

Semantics-based Graph Approach to Complex Question-Answering
Tomasz Jurczyk and Jinho D. Choi

This paper suggests an architectural approach of representing knowledge graph for complex question-answering. There are four kinds of entity relations added to our knowledge graph: syntactic dependencies, semantic role labels, named entities, and coreference links, which can be effectively applied to answer complex questions. As a proof of concept, we demonstrate how our knowledge graph can be used to solve complex questions such as arithmetics. Our experiment shows a promising result on solving arithmetic questions, achieving the 3-folds cross-validation score of 71.75%.

Recognizing Textual Entailment using Dependency Analysis and Machine Learning
Nidhi Sharma, Richa Sharma, and Kanad K. Biswas

This paper presents a machine learning system that uses dependency-based features and lexical features for recognizing textual entailment. The system first generates a frame-based structured representation of the sentences using dependency relations obtained from Stanford Dependency parser. This structured representation is then used to evaluate a set of lexical and syntactic features. These features are simple and intuitive; easy to comprehend and evaluate. The system evaluates the feature values automatically without any manual effort or intervention. The performance of the proposed system is evaluated by performing experiments on RTE1, RTE2 and RTE3 datasets. Decision trees are found to outperform other classifiers while SVM classifier using an RBF kernel shows comparable performance. Additionally, a comparative study of the current system with other ML-based systems for RTE to check the performance of the proposed system is also carried out. The dependency-based heuristics and lexical features from the current system have resulted in significant improvement in accuracy over existing state-of-art ML-based solutions for RTE. A task-based analysis of the current system for RTE1 dataset is also performed which, shows improved performance as compared to similar study performed by different researchers in the past.

Bilingual lexicon extraction for a distant language pair using a small parallel corpus
Ximena Gutierrez-Vasques

The aim of this thesis proposal is to perform bilingual lexicon extraction for cases in which small parallel corpora are available and it is not easy to obtain monolingual corpus for at least one of the languages. Moreover, the languages are typologically distant and there is no bilingual seed lexicon available. We focus on the language pair Spanish-Nahuatl, we propose to work with morpheme based representations in order to reduce the sparseness and to facilitate the task of finding lexical correspondences between a highly agglutinative language and a fusional one. We take into account contextual information but instead of using a precompiled seed dictionary, we use the distribution and dispersion of the positions of the morphological units as cues to compare the contextual vectors and obtaining the translation candidates.

Morphological Paradigms: Computational Structure and Unsupervised Learning
Jackson Lee

This thesis explores the computational structure of morphological paradigms from the perspective of unsupervised learning. Three topics are studied: (i) stem identification, (ii) paradigmatic similarity, and (iii) paradigm induction. All the three topics progress in terms of the scope of data in question. The first and second topics explore structure when morphological paradigms are given, first within a paradigm and then across paradigms. The third topic asks where morphological paradigms come from in the first place, and explores strategies of paradigm induction from child-directed speech. This research is of interest to linguists and natural language processing researchers, for both theoretical questions and applied areas.

Computational Exploration to Linguistic Structures of Future: Classification and Categorization
Aiming Ni, Jinho D. Choi, Jason Shepard, and Phillip Wolff

Many languages, like English, lack a future tense, which makes the automatic identification of future reference a challenge. In this research we extend Latent Dirichlet allocation (LDA) for use in the identification of future referring sentences. Building off a set of hand designed rules, we trained a ADAGRAD classifier to be able automatically detect sentences referring to the future. Uni-bi-trigram and syntactic rule mixed feature was found to provide the highest accuracy. Latent Dirichlet Allocation (LDA) indicated the existence of four major categories future orientation. Lastly, the results of these analyses were found to correlate with a range of behavioral measures, offering evidence in support of the psychological reality of the categories.

4

Main Conference: Tuesday, June 2

Overview

7:30 – 9:00	**Registration and Breakfast**		*Plaza Exhibit All*
	Session 4		
9:00 – 10:40	Dialogue and Spoken Language Processing (Long Papers) *Plaza Ballroom A & B*	Machine Learning for NLP (Long Papers) *Plaza Ballroom D & E*	Phonology, Morphology and Word Segmentation (Long Papers) *Plaza Ballroom F*
10:40 – 11:15	**Break**		*Plaza Exhibit All*
	Session 5		
11:15 – 12:30	Semantics (Short Papers) *Plaza Ballroom A & B*	Machine Translation (Short Papers) *Plaza Ballroom D & E*	Morphology, Syntax, Multilinguality, and Applications (Short Papers) *Plaza Ballroom F*
12:30 – 2:00	**Lunch**		
	Session 6		
2:00 – 3:15	Generation and Summarization (Long Papers) *Plaza Ballroom A & B*	Discourse and Coreference (Long Papers) *Plaza Ballroom D & E*	Information Extraction and Question Answering (Long Papers) *Plaza Ballroom F*
3:15 – 3:45	**Break**		*Plaza Exhibit All*
	Session 7		
3:45 – 5:00	Semantics (Long + TACL Papers) *Plaza Ballroom A & B*	Information Extraction and Question Answering (Long + TACL Papers) *Plaza Ballroom D & E*	Machine Translation (Long Papers) *Plaza Ballroom F*
5:00 – 6:30	**Poster session 2A (Short papers)**		*Plaza Ballroom A, B, & C*
5:00 – 6:30	**Demo Session A**		*Plaza Ballroom A, B, & C*
6:30 – 8:00	**Poster session 2B (Short papers)**		*Plaza Ballroom A, B, & C*

6:30 – 8:00	**Demo Session B**	*Plaza Ballroom A, B, & C*
8:00 – 11:00	**Social Event**	*Grand Ballroom*

Session 4 Overview – Tuesday, June 2, 2015

Track A	Track B	Track C	
Dialogue and Spoken Language Processing (Long Papers) Plaza Ballroom A & B	*Machine Learning for NLP (Long Papers)* Plaza Ballroom D & E	*Phonology, Morphology and Word Segmentation (Long Papers)* Plaza Ballroom F	
Semantic Grounding in Dialogue for Complex Problem Solving *Li and Boyer*	Early Gains Matter: A Case for Preferring Generative over Discriminative Crowdsourcing Models *Felt, Black, Ringger, Seppi, and Haertel*	Inflection Generation as Discriminative String Transduction *Nicolai, Cherry, and Kondrak*	9:00
Learning Knowledge Graphs for Question Answering through Conversational Dialog *Hixon, Clark, and Hajishirzi*	Optimizing Multivariate Performance Measures for Learning Relation Extraction Models *Haffari, Nagesh, and Ramakrishnan*	Penalized Expectation Propagation for Graphical Models over Strings *Cotterell and Eisner*	9:25
Sentence segmentation of aphasic speech *Fraser, Ben-David, Hirst, Graham, and Rochon*	Convolutional Neural Network for Paraphrase Identification *Yin and Schütze*	Joint Generation of Transliterations from Multiple Representations *Yao and Kondrak*	9:50
Semantic parsing of speech using grammars learned with weak supervision *Gaspers, Cimiano, and Wrede*	Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval *Liu, Gao, He, Deng, Duh, and Wang*	Prosodic boundary information helps unsupervised word segmentation *Ludusan, Synnaeve, and Dupoux*	10:15

Parallel Session 4

Session 4A: Dialogue and Spoken Language Processing (Long Papers)

Plaza Ballroom A & B *Chair: Marilyn Walker*

Semantic Grounding in Dialogue for Complex Problem Solving

Xiaolong Li and Kristy Boyer 09:00–09:25

Dialogue systems that support users in complex problem solving must interpret user utterances within the context of a dynamically changing, user-created problem solving artifact. This paper presents a novel approach to semantic grounding of noun phrases within tutorial dialogue for computer programming. Our approach performs joint segmentation and labeling of the noun phrases to link them to attributes of entities within the problem-solving environment. Evaluation results on a corpus of tutorial dialogue for Java programming demonstrate that a Conditional Random Field model performs well, achieving an accuracy of 89.3% for linking semantic segments to the correct entity attributes. This work is a step toward enabling dialogue systems to support users in increasingly complex problem-solving tasks.

Learning Knowledge Graphs for Question Answering through Conversational Dialog

Ben Hixon, Peter Clark, and Hannaneh Hajishirzi 09:25–09:50

We describe how a question-answering system can learn about its domain from conversational dialogs. Our system learns to relate concepts in science questions to propositions in a fact corpus, stores new concepts and relations in a knowledge graph (KG), and uses the graph to solve questions. We are the first to acquire knowledge for question-answering from open, natural language dialogs without a fixed ontology or domain model that predetermines what users can say. Our relation-based strategies complete more successful dialogs than a query expansion baseline, our task-driven relations are more effective for solving science questions than relations from general knowledge sources, and our method is practical enough to generalize to other domains.

Sentence segmentation of aphasic speech

Kathleen C. Fraser, Naama Ben-David, Graeme Hirst, Naida Graham, and Elizabeth Rochon 09:50–10:15

Automatic analysis of impaired speech for screening or diagnosis is a growing research field; however there are still many barriers to a fully automated approach. When automatic speech recognition is used to obtain the speech transcripts, sentence boundaries must be inserted before most measures of syntactic complexity can be computed. In this paper, we consider how language impairments can affect segmentation methods, and compare the results of computing syntactic complexity metrics on automatically and manually segmented transcripts. We find that the important boundary indicators and the resulting segmentation accuracy can vary depending on the type of impairment observed, but that results on patient data are generally similar to control data. We also find that a number of syntactic complexity metrics are robust to the types of segmentation errors that are typically made.

Semantic parsing of speech using grammars learned with weak supervision

Judith Gaspers, Philipp Cimiano, and Britta Wrede 10:15–10:40

Semantic grammars can be applied both as a language model for a speech recognizer and for semantic parsing, e.g. in order to map the output of a speech recognizer into formal meaning representations. Semantic speech recognition grammars are, however, typically created manually or learned in a supervised fashion, requiring extensive manual effort in both cases. Aiming to reduce this effort, in this paper we investigate the induction of semantic speech recognition grammars under weak supervision. We present empirical results, indicating that the induced grammars support semantic parsing of speech with a rather low loss in performance when compared to parsing of input without recognition errors. Further, we show improved parsing performance compared to applying n-gram models as language models and demonstrate how our semantic speech recognition grammars can be enhanced by weights based on occurrence frequencies, yielding an improvement in parsing performance over applying unweighted grammars.

Session 4B: Machine Learning for NLP (Long Papers)

Plaza Ballroom D & E *Chair: Amarnag Subramanya*

Early Gains Matter: A Case for Preferring Generative over Discriminative Crowdsourcing Models

Paul Felt, Kevin Black, Eric Ringger, Kevin Seppi, and Robbie Haertel 09:00–09:25

In modern practice, labeling a dataset often involves aggregating annotator judgments obtained from crowdsourcing. State-of-the-art aggregation is performed via inference on probabilistic models, some of which are data-aware, meaning that they leverage features of the data (e.g., words in a document) in addition to annotator judgments. Previous work largely prefers discriminatively trained conditional models. This paper demonstrates that a data-aware crowdsourcing model incorporating a generative multinomial data model enjoys a strong competitive advantage over its discriminative log-linear counterpart in the typical crowdsourcing setting. That is, the generative approach is better except when the annotators are highly accurate in which case simple majority vote is often sufficient. Additionally, we present a novel mean-field variational inference algorithm for the generative model that significantly improves on the previously reported state-of-the-art for that model. We validate our conclusions on six text classification datasets with both human-generated and synthetic annotations.

Optimizing Multivariate Performance Measures for Learning Relation Extraction Models

Gholamreza Haffari, Ajay Nagesh, and Ganesh Ramakrishnan 09:25–09:50

We describe a novel max-margin learning approach to optimize non-linear performance measures for distantly-supervised relation extraction models. Our approach can be generally used to learn latent variable models under multivariate non-linear performance measures, such as F_β-score. Our approach interleaves Concave-Convex Procedure (CCCP) for populating latent variables with dual decomposition to factorize the original hard problem into smaller independent sub-problems. The experimental results demonstrate that our learning algorithm is more effective than the ones commonly used in the literature for distant supervision of information extraction models. On several data conditions, we show that our method outperforms the baseline and results in up to 8.5% improvement in the F_1-score.

Convolutional Neural Network for Paraphrase Identification

Wenpeng Yin and Hinrich Schütze 09:50–10:15

We present a new deep learning architecture Bi-CNN-MI for paraphrase identification (PI). Based on the insight that PI requires comparing two sentences on multiple levels of granularity, we learn multigranular sentence representations using convolutional neural network (CNN) and model interaction features at each level. These features are then the input to a logistic classifier for PI. All parameters of the model (for embeddings, convolution and classification) are directly optimized for PI. To address the lack of training data, we pretrain the network in a novel way using a language modeling task. Results on the MSRP corpus surpass that of previous NN competitors.

Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval

Xiaodong Liu, Jianfeng Gao, Xiaodong He, Li Deng, Kevin Duh, and Ye-Yi Wang 10:15–10:40

Methods of deep neural networks (DNNs) have recently demonstrated superior performance on a number of natural language processing tasks. However, in most previous work, the models are learned based on either unsupervised objectives, which does not directly optimize the desired task, or single- task supervised objectives, which often suffer from insufficient training data. We develop a multi-task DNN for learning representations across multiple tasks, not only leveraging large amounts of cross-task data, but also benefiting from a regularization effect that leads to more general representations to help tasks in new domains. Our multi-task DNN approach combines tasks of multiple-domain classification (for query classification) and information retrieval (ranking for web search), and demonstrates significant gains over strong baselines in a comprehensive set of domain adaptation and other multi-task learning experiments.

Session 4C: Phonology, Morphology and Word Segmentation (Long Papers)

Plaza Ballroom F *Chair: Hal Daumé III*

Inflection Generation as Discriminative String Transduction
Garrett Nicolai, Colin Cherry, and Grzegorz Kondrak 09:00–09:25

We approach the task of morphological inflection generation as discriminative string transduction. Our supervised system learns to generate word-forms from lemmas accompanied by morphological tags, and refines them by referring to the other forms within a paradigm. Results of experiments on six diverse languages with varying amounts of training data demonstrate that our approach improves the state of the art in terms of predicting inflected word-forms.

Penalized Expectation Propagation for Graphical Models over Strings
Ryan Cotterell and Jason Eisner 09:25–09:50

We present penalized expectation propagation, a novel algorithm for approximate inference in graphical models. Expectation propagation is a variant of loopy belief propagation that keeps messages tractable by projecting them back into a given family of functions. Our extension speeds up the method by using a structured-sparsity penalty to prefer simpler messages within the family. In the case of string-valued random variables, penalized EP lets us work with an expressive non-parametric function family based on variable-length n-gram models. On phonological inference problems, we obtain substantial speedup over previous related algorithms with no significant loss in accuracy.

Joint Generation of Transliterations from Multiple Representations
Lei Yao and Grzegorz Kondrak 09:50–10:15

Machine transliteration is often referred to as phonetic translation. We show that transliterations incorporate information from both spelling and pronunciation, and propose an effective model for joint transliteration generation from both representations. We further generalize this model to include transliterations from other languages, and enhance it with re-ranking and lexicon features. We demonstrate substantial improvements in transliteration accuracy on several datasets.

Prosodic boundary information helps unsupervised word segmentation
Bogdan Ludusan, Gabriel Synnaeve, and Emmanuel Dupoux 10:15–10:40

It is well known that prosodic information is used by infants in early language acquisition. In particular, prosodic boundaries have been shown to help infants with sentence and word-level segmentation. In this study, we extend an unsupervised method for word segmentation to include information about prosodic boundaries. The boundary information used was either derived from oracle data (hand-annotated), or extracted automatically with a system that employs only acoustic cues for boundary detection. The approach was tested on two different languages, English and Japanese, and the results show that boundary information helps word segmentation in both cases. The performance gain obtained for two typologically distinct languages shows the robustness of prosodic information for word segmentation. Furthermore, the improvements are not limited to the use of oracle information, similar performances being obtained also with automatically extracted boundaries.

Session 5 Overview – Tuesday, June 2, 2015

Track A	Track B	Track C	
Semantics (Short Papers) Plaza Ballroom A & B	*Machine Translation (Short Papers)* Plaza Ballroom D & E	*Morphology, Syntax, Multilinguality, and Applications (Short Papers)* Plaza Ballroom F	
So similar and yet incompatible: Toward the automated identification of semantically compatible words *Kruszewski and Baroni*	Morphological Modeling for Machine Translation of English-Iraqi Arabic Spoken Dialogs *Kirchhoff, Tam, Richey, and Wang*	Paradigm classification in supervised learning of morphology *Ahlberg, Forsberg, and Hulden*	11:15
Do Supervised Distributional Methods Really Learn Lexical Inference Relations? *Levy, Remus, Biemann, and Dagan*	Continuous Adaptation to User Feedback for Statistical Machine Translation *Blain, Bougares, Hazem, Barrault, and Schwenk*	Shift-Reduce Constituency Parsing with Dynamic Programming and POS Tag Lattice *Mi and Huang*	11:30
A Word Embedding Approach to Predicting the Compositionality of Multiword Expressions *Salehi, Cook, and Baldwin*	Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation *Xing, Wang, Liu, and Lin*	Unsupervised Code-Switching for Multilingual Historical Document Transcription *Garrette, Alpert-Abrams, Berg-Kirkpatrick, and Klein*	11:45
Word Embedding-based Antonym Detection using Thesauri and Distributional Information *Ono, Miwa, and Sasaki*	Fast and Accurate Preordering for SMT using Neural Networks *de Gispert, Iglesias, and Byrne*	Matching Citation Text and Cited Spans in Biomedical Literature: a Search-Oriented Approach *Cohan, Soldaini, and Goharian*	12:00
A Comparison of Word Similarity Performance Using Explanatory and Non-explanatory Texts *Jin and Schuler*	APRO: All-Pairs Ranking Optimization for MT Tuning *Dreyer and Dong*	Effective Feature Integration for Automated Short Answer Scoring *Sakaguchi, Heilman, and Madnani*	12:15

Parallel Session 5

Session 5A: Semantics (Short Papers)

Plaza Ballroom A & B — *Chair: Daniel Gildea*

So similar and yet incompatible: Toward the automated identification of semantically compatible words
Germán Kruszewski and Marco Baroni 11:15–11:30

We introduce the challenge of detecting semantically compatible words, that is, words that can potentially refer to the same thing (cat and hindrance are compatible, cat and dog are not), arguing for its central role in many semantic tasks. We present a publicly available data-set of human compatibility ratings, and a neural-network model that takes distributional embeddings of words as input and learns alternative embeddings that perform the compatibility detection task quite well.

Do Supervised Distributional Methods Really Learn Lexical Inference Relations?
Omer Levy, Steffen Remus, Chris Biemann, and Ido Dagan 11:30–11:45

Distributional representations of words have been recently used in supervised settings for recognizing lexical inference relations between word pairs, such as hypernymy and entailment. We investigate a collection of these state-of-the-art methods, and show that they do not actually learn a relation between two words. Instead, they learn an independent property of a single word in the pair: whether that word is a "prototypical hypernym".

A Word Embedding Approach to Predicting the Compositionality of Multiword Expressions
Bahar Salehi, Paul Cook, and Timothy Baldwin 11:45–12:00

This paper presents the first attempt to use word embeddings to predict the compositionality of multiword expressions. We consider both single- and multi-prototype word embeddings. Experimental results show that, in combination with a back-off method based on string similarity, word embeddings outperform a method using count-based distributional similarity. Our best results are competitive with, or superior to, state-of-the-art methods over three standard compositionality datasets, which include two types of multiword expressions and two languages.

Word Embedding-based Antonym Detection using Thesauri and Distributional Information
Masataka Ono, Makoto Miwa, and Yutaka Sasaki 12:00–12:15

This paper proposes a novel approach to train word embeddings to capture antonyms. Word embeddings have shown to capture synonyms and analogies. Such word embeddings, however, cannot capture antonyms since they depend on the distributional hypothesis. Our approach utilizes supervised synonym and antonym information from thesauri, as well as distributional information from large-scale unlabelled text data. The evaluation results on the GRE antonym question task show that our model outperforms the state-of-the-art systems and it can answer the antonym questions in the F-score of 89%.

A Comparison of Word Similarity Performance Using Explanatory and Non-explanatory Texts
Lifeng Jin and William Schuler 12:15–12:30

Vectorial representations derived from large current events datasets such as Google News have been shown to perform well on word similarity tasks. This paper shows vectorial representations derived from substantially smaller explanatory text datasets such as English Wikipedia and Simple English Wikipedia preserve enough lexical semantic information to make these kinds of category judgments with equal or better accuracy. Analysis shows these results are driven by a prevalence of commonsense facts in explanatory text. These positive results for small datasets suggest vectors derived from slower but more accurate deep parsers may be practical for lexical semantic applications.

Session 5B: Machine Translation (Short Papers)

Plaza Ballroom D & E *Chair: David Chiang*

Morphological Modeling for Machine Translation of English-Iraqi Arabic Spoken Dialogs
Katrin Kirchhoff, Yik-Cheung Tam, Colleen Richey, and Wen Wang 11:15–11:30

This paper addresses the problem of morphological modeling in statistical speech-to-speech translation for English to Iraqi Arabic. An analysis of user data from a real-time MT-based dialog system showed that generating correct verbal inflections is a key problem for this language pair. We approach this problem by enriching the training data with morphological information derived from source-side dependency parses. We analyze the performance of several parsers as well as the effect on different types of translation models. Our method achieves an improvement of more than a full BLEU point and a significant increase in verbal inflection accuracy; at the same time, it is computationally inexpensive and does not rely on target-language linguistic tools.

Continuous Adaptation to User Feedback for Statistical Machine Translation
Frédéric Blain, Fethi Bougares, Amir Hazem, Loïc Barrault, and Holger Schwenk 11:30–11:45

This paper gives a detailed experiment feedback of different approaches to adapt a statistical machine translation system towards a targeted translation project, using only small amounts of parallel in-domain data. The experiments were performed by professional translators under realistic conditions of work using a computer assisted translation tool. We analyze the influence of these adaptations on the translator productivity and on the overall post-editing effort. We show that significant improvements can be obtained by using the presented adaptation techniques.

Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation
Chao Xing, Dong Wang, Chao Liu, and Yiye Lin 11:45–12:00

Word embedding has been found to be highly powerful to translate words from one language to another by a simple linear transform. However, we found some inconsistence among the objective functions of the embedding and the transform learning, as well as the distance measuring. This paper proposes a solution which normalizes the word vectors on a hypersphere and constrains the linear transform as a orthogonal transform. The experimental results confirmed that the proposed solution can offer better performance on a word similarity task and an English-to-Spanish word translation task.

Fast and Accurate Preordering for SMT using Neural Networks
Adrià de Gispert, Gonzalo Iglesias, and Bill Byrne 12:00–12:15

We propose the use of neural networks to model source-side preordering for faster and better statistical machine translation. The neural network trains a logistic regression model to predict whether two sibling nodes of the source-side parse tree should be swapped in order to obtain a more monotonic parallel corpus, based on samples extracted from the word-aligned parallel corpus. For multiple language pairs and domains, we show that this yields the best reordering performance against other state-of-the-art techniques, resulting in improved translation quality and very fast decoding.

APRO: All-Pairs Ranking Optimization for MT Tuning
Markus Dreyer and Yuanzhe Dong 12:15–12:30

We present APRO, a new method for machine translation tuning that can handle large feature sets. As opposed to other popular methods (e.g., MERT, MIRA, PRO), which involve randomness and require multiple runs to obtain a reliable result, APRO gives the same result on any run, given initial feature weights. APRO follows the pairwise ranking approach of PRO (Hopkins and May, 2011), but instead of ranking a small sampled subset of pairs from the k- best list, APRO efficiently ranks all pairs. By obviating the need for manually determined sampling settings, we obtain more reliable results. APRO converges more quickly than PRO and gives similar or better translation results.

Session 5C: Morphology, Syntax, Multilinguality, and Applications (Short Papers)

Plaza Ballroom F *Chair: William Lewis*

Paradigm classification in supervised learning of morphology

Malin Ahlberg, Markus Forsberg, and Mans Hulden 11:15–11:30

Supervised morphological paradigm learning by identifying and aligning the longest common subsequence found in inflection tables has recently been proposed as a simple yet competitive way to induce morphological patterns. We combine this non-probabilistic strategy of inflection table generalization with a discriminative classifier to permit the reconstruction of complete inflection tables of unseen words. Our system learns morphological paradigms from labeled examples of inflection patterns (inflection tables) and then produces inflection tables from unseen lemmas or base forms. We evaluate the approach on datasets covering 11 different languages and show that this approach results in consistently higher accuracies vis-à-vis other methods on the same task, thus indicating that the general method is a viable approach to quickly creating high-accuracy morphological resources.

Shift-Reduce Constituency Parsing with Dynamic Programming and POS Tag Lattice

Haitao Mi and Liang Huang 11:30–11:45

We present the first dynamic programming (DP) algorithm for shift-reduce constituency parsing, which extends the DP idea of Huang and Sagae (2010) to context-free grammars. To alleviate the propagation of errors from part-of-speech tagging, we also extend the parser to take a tag lattice instead of a fixed tag sequence. Experiments on both English and Chinese treebanks show that our DP parser significantly improves parsing quality over non-DP baselines, and achieves the best accuracies among empirical linear-time parsers.

Unsupervised Code-Switching for Multilingual Historical Document Transcription

Dan Garrette, Hannah Alpert-Abrams, Taylor Berg-Kirkpatrick, and Dan Klein 11:45–12:00

Transcribing documents from the printing press era, a challenge in its own right, is more complicated when documents interleave multiple languages—a common feature of 16th century texts. Additionally, many of these documents precede consistent orthographic conventions, making the task even harder. We extend the state-of-the-art historical OCR model of Berg-Kirkpatrick et al. (2013) to handle word-level code-switching between multiple languages. Further, we enable our system to handle spelling variability, including now-obsolete shorthand systems used by printers. Our results show average relative character error reductions of 14% across a variety of historical texts.

Matching Citation Text and Cited Spans in Biomedical Literature: a Search-Oriented Approach

Arman Cohan, Luca Soldaini, and Nazli Goharian 12:00–12:15

Citation sentences (citances) to a reference ar- ticle have been extensively studied for sum- marization tasks. However, citances might not accurately represent the content of the cited article, as they often fail to capture the con- text of the reported findings and can be af- fected by epistemic value drift. Following the intuition behind the TAC (Text Analysis Conference) 2014 Biomedical Summarization track, we propose a system that identifies text spans in the reference article that are related to a given citance. We refer to this problem as citance-reference spans matching. We ap- proach the problem as a retrieval task; in this paper, we detail a comparison of different ci- tance reformulation methods and their combi- nations. While our results show improvement over the baseline (up to 25.9%), their absolute magnitude implies that there is ample room for future improvement.

Effective Feature Integration for Automated Short Answer Scoring

Keisuke Sakaguchi, Michael Heilman, and Nitin Madnani 12:15–12:30

A major opportunity for NLP to have a real-world impact is in helping educators score student writing, particularly content-based writing (i.e., the task of automated short answer scoring). A major challenge in this enterprise is that scored responses to a particular question (i.e., labeled data) are valuable for modeling but limited in quantity. Additional information from the scoring guidelines for humans, such as exemplars for each score level and descriptions of key concepts, can also be used. Here, we explore methods for integrating scoring guidelines and labeled responses, and we find that stacked generalization (Wolpert, 1992) improves performance, especially for small training sets.

Session 6 Overview – Tuesday, June 2, 2015

Track A	**Track B**	**Track C**	
Generation and Summarization (Long Papers) Plaza Ballroom A & B	*Discourse and Coreference (Long Papers)* Plaza Ballroom D & E	*Information Extraction and Question Answering (Long Papers)* Plaza Ballroom F	
Socially-Informed Timeline Generation for Complex Events *Wang, Cardie, and Marchetti*	Encoding World Knowledge in the Evaluation of Local Coherence *Zhang, Feng, Qin, Hirst, Liu, and Huang*	Injecting Logical Background Knowledge into Embeddings for Relation Extraction *Rocktäschel, Singh, and Riedel*	2:00
Movie Script Summarization as Graph-based Scene Extraction *Gorinski and Lapata*	Chinese Event Coreference Resolution: An Unsupervised Probabilistic Model Rivaling Supervised Resolvers *Chen and Ng*	Unsupervised Entity Linking with Abstract Meaning Representation *Pan, Cassidy, Hermjakob, Ji, and Knight*	2:25
Toward Abstractive Summarization Using Semantic Representations *Liu, Flanigan, Thomson, Sadeh, and Smith*	Removing the Training Wheels: A Coreference Dataset that Entertains Humans and Challenges Computers *Guha, Iyyer, Bouman, and Boyd-Graber*	Idest: Learning a Distributed Representation for Event Patterns *Krause, Alfonseca, Filippova, and Pighin*	2:50

Parallel Session 6

Session 6A: Generation and Summarization (Long Papers)

Plaza Ballroom A & B *Chair: Wei Xu*

Socially-Informed Timeline Generation for Complex Events
Lu Wang, Claire Cardie, and Galen Marchetti 14:00–14:25

Existing timeline generation systems for complex events consider only information from traditional media, ignoring the rich social context provided by user-generated content that reveals representative public interests or insightful opinions. We instead aim to generate socially-informed timelines that contain both news article summaries and selected user comments. We present an optimization framework designed to balance topical cohesion between the article and comment summaries along with their informativeness and coverage of the event. Automatic evaluations on real-world datasets that cover four complex events show that our system produces more informative timelines than state-of-the-art systems. In human evaluation, the associated comment summaries are furthermore rated more insightful than editor's picks and comments ranked highly by users.

Movie Script Summarization as Graph-based Scene Extraction
Philip John Gorinski and Mirella Lapata 14:25–14:50

In this paper we study the task of movie script summarization, which we argue could enhance script browsing, give readers a rough idea of the script's plotline, and speed up reading time. We formalize the process of generating a shorter version of a screenplay as the task of finding an optimal chain of scenes. We develop a graph-based model that selects a chain by jointly optimizing its logical progression, diversity, and importance. Human evaluation based on a question-answering task shows that our model produces summaries which are more informative compared to competitive baselines.

Toward Abstractive Summarization Using Semantic Representations
Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A. Smith 14:50–15:15

We present a novel abstractive summarization framework that draws on the recent development of a treebank for the Abstract Meaning Representation (AMR). In this framework, the source text is parsed to a set of AMR graphs, the graphs are transformed into a summary graph, and then text is generated from the summary graph. We focus on the graph-to-graph transformation that reduces the source semantic graph into a summary graph, making use of an existing AMR parser and assuming the eventual availability of an AMR-to-text generator. The framework is data-driven, trainable, and not specifically designed for a particular domain. Experiments on gold-standard AMR annotations and system parses show promising results. Code is available at: https://github.com/summarization

Session 6B: Discourse and Coreference (Long Papers)

Plaza Ballroom D & E *Chair: Ani Nenkova*

Encoding World Knowledge in the Evaluation of Local Coherence
Muyu Zhang, Vanessa Wei Feng, Bing Qin, Graeme Hirst, Ting Liu, and Jingwen Huang 14:00–14:25

Previous work on text coherence was primarily based on matching multiple mentions of the same entity in different parts of the text; therefore, it misses the contribution from semantically related but not necessarily coreferential entities (e.g., Gates and Microsoft). In this paper, we capture such semantic relatedness by leveraging world knowledge (e.g., Gates is the person who created Microsoft), and use two existing evaluation frameworks. First, in the unsupervised framework, we introduce semantic relatedness as an enrichment to the original graph-based model of Guinaudeau and Strube (2013). In addition, we incorporate semantic relatedness as additional features into the popular entity-based model of Barzilay and Lapata (2008). Across both frameworks, our enriched model with semantic relatedness outperforms the original methods, especially on short documents.

Chinese Event Coreference Resolution: An Unsupervised Probabilistic Model Rivaling Supervised Resolvers
Chen Chen and Vincent Ng 14:25–14:50

Recent work has successfully leveraged the semantic information extracted from lexical knowledge bases such as WordNet and FrameNet to improve English event coreference resolvers. The lack of comparable resources in other languages, however, has made the design of high-performance non-English event coreference resolvers, particularly those employing unsupervised models, very difficult. We propose a generative model for the under-studied task of Chinese event coreference resolution that rivals its supervised counterparts in performance when evaluated on the ACE 2005 corpus.

Removing the Training Wheels: A Coreference Dataset that Entertains Humans and Challenges Computers
Anupam Guha, Mohit Iyyer, Danny Bouman, and Jordan Boyd-Graber 14:50–15:15

Coreference is a core NLP problem. However, newswire data, the primary source of existing coreference data, lack the richness necessary to truly solve coreference. We present a new domain with denser references—quiz bowl questions—that is challenging and enjoyable to humans, and we use the quiz bowl community to develop a new coreference dataset, together with an annotation framework that can tag any text data with coreferences and named entities. We also successfully integrate active learning into this annotation pipeline to collect documents maximally useful to coreference models. State-of-the-art coreference systems underperform a simple classifier on our new dataset, motivating non-newswire data for future coreference research.

Session 6C: Information Extraction and Question Answering (Long Papers)

Plaza Ballroom F *Chair: Lucy Vanderwende*

Injecting Logical Background Knowledge into Embeddings for Relation Extraction

Tim Rocktäschel, Sameer Singh, and Sebastian Riedel 14:00–14:25

Matrix factorization approaches to relation extraction provide several attractive features: they support distant supervision, handle open schemas, and leverage unlabeled data. Unfortunately, these methods share a shortcoming with all other distantly supervised approaches: they cannot learn to extract target relations without existing data in the knowledge base, and likewise, these models are inaccurate for relations with sparse data. Rule-based extractors, on the other hand, can be easily extended to novel relations and improved for existing but inaccurate relations, through first-order formulae that capture auxiliary domain knowledge. However, usually a large set of such formulae is necessary to achieve generalization. In this paper, we introduce a paradigm for learning low-dimensional embeddings of entity-pairs and relations that combine the advantages of matrix factorization with first-order logic domain knowledge. We introduce simple approaches for estimating such embeddings, as well as a novel training algorithm to jointly optimize over factual and first-order logic information. Our results show that this method is able to learn accurate extractors with little or no distant supervision alignments, while at the same time generalizing to textual patterns that do not appear in the formulae.

Unsupervised Entity Linking with Abstract Meaning Representation

Xiaoman Pan, Taylor Cassidy, Ulf Hermjakob, Heng Ji, and Kevin Knight 14:25–14:50

Most successful Entity Linking (EL) methods aim to link mentions to their referent entities in a structured Knowledge Base (KB) by comparing their respective contexts, often using similarity measures. While the KB structure is given, current methods have suffered from impoverished information representations on the mention side. In this paper, we demonstrate the effectiveness of Abstract Meaning Representation (AMR) (Banarescu et al., 2013) to select high quality sets of entity "collaborators" to feed a simple similarity measure (Jaccard) to link entity mentions. Experimental results show that AMR captures contextual properties discriminative enough to make linking decisions, without the need for EL training data, and that system with AMR parsing output outperforms hand labeled traditional semantic roles as context representation for EL. Finally, we show promising preliminary results for using AMR to select sets of "coherent" entity mentions for collective entity linking.

Idest: Learning a Distributed Representation for Event Patterns

Sebastian Krause, Enrique Alfonseca, Katja Filippova, and Daniele Pighin 14:50–15:15

This paper describes Idest, a new method for learning paraphrases of event patterns. It is based on a new neural network architecture that only relies on the weak supervision signal that comes from the news published on the same day and mention the same real-world entities. It can generalize across extractions from different dates to produce a robust paraphrase model for event patterns that can also capture meaningful representations for rare patterns. We compare it with two state-of-the-art systems and show that it can attain comparable quality when trained on a small dataset. Its generalization capabilities also allow it to leverage much more data, leading to substantial quality improvements.

Session 7 Overview – Tuesday, June 2, 2015

Track A	Track B	Track C
Semantics (Long + TACL Papers) Plaza Ballroom A & B	*Information Extraction and Question Answering (Long + TACL Papers)* Plaza Ballroom D & E	*Machine Translation (Long Papers)* Plaza Ballroom F
High-Order Low-Rank Tensors for Semantic Role Labeling *Lei, Zhang, Màrquez, Moschitti, and Barzilay*	Lexical Event Ordering with an Edge-Factored Model *Abend, Cohen, and Steedman*	Bag-of-Words Forced Decoding for Cross-Lingual Information Retrieval *Hieber and Riezler*
[TACL] Large-scale Semantic Parsing without Question-Answer Pairs *Reddy, Lapata, and Steedman*	[TACL] Entity disambiguation with web links *Chisholm and Hachey*	Accurate Evaluation of Segment-level Machine Translation Metrics *Graham, Baldwin, and Mathur*
[TACL] A Large Scale Evaluation of Distributional Semantic Models: Parameters, Interactions and Model Selection *Lapesa and Evert*	[TACL] A Joint Model for Entity Analysis: Coreference, Typing, and Linking *Durrett and Klein*	Leveraging Small Multilingual Corpora for SMT Using Many Pivot Languages *Dabre, Cromieres, Kurohashi, and Bhattacharyya*

3:45
4:10
4:35

Parallel Session 7

Session 7A: Semantics (Long + TACL Papers)

Plaza Ballroom A & B *Chair: Xiaodong He*

High-Order Low-Rank Tensors for Semantic Role Labeling
Tao Lei, Yuan Zhang, Lluís Màrquez, Alessandro Moschitti, and Regina Barzilay 15:45–16:10

This paper introduces a tensor-based approach to semantic role labeling (SRL). The motivation behind the approach is to automatically induce a compact feature representation for words and their relations, tailoring them to the task. In this sense, our dimensionality reduction method provides a clear alternative to the traditional feature engineering approach used in SRL. To capture meaningful interactions between the argument, predicate, their syntactic path and the corresponding role label, we compress each feature representation first to a lower dimensional space prior to assessing their interactions. This corresponds to using an overall cross-product feature representation and maintaining associated parameters as a four-way low-rank tensor. The tensor parameters are optimized for the SRL performance using standard online algorithms. Our tensor-based approach rivals the best performing system on the CoNLL-2009 shared task. In addition, we demonstrate that adding the representation tensor to a competitive tensor-free model yields 2% absolute increase in F-score.

[TACL] Large-scale Semantic Parsing without Question-Answer Pairs
Siva Reddy, Mirella Lapata, and Mark Steedman 16:10–16:35

In this paper we introduce a novel semantic parsing approach to query Freebase in natural language without requiring manual annotations or question-answer pairs. Our key insight is to represent natural language via semantic graphs whose topology shares many commonalities with Freebase. Given this representation, we conceptualize semantic parsing as a graph matching problem. Our model converts sentences to semantic graphs using CCG and subsequently grounds them to Freebase guided by denotations as a form of weak supervision. Evaluation experiments on a subset of the Free917 and WebQuestions benchmark datasets show our semantic parser improves over the state of the art.

[TACL] A Large Scale Evaluation of Distributional Semantic Models: Parameters, Interactions and Model Selection
Gabriella Lapesa and Stefan Evert 16:35–17:00

This paper presents the results of a large-scale evaluation study of window-based Distributional Semantic Models on a wide variety of tasks. Our study combines a broad coverage of model parameters with a model selection methodology that is robust to overfitting and able to capture parameter interactions. We show that our strategy allows us to identify parameter configurations that achieve good performance across different datasets and tasks.

Session 7B: Information Extraction and Question Answering (Long + TACL Papers)

Plaza Ballroom D & E *Chair: Vincent Ng*

Lexical Event Ordering with an Edge-Factored Model
Omri Abend, Shay B. Cohen, and Mark Steedman 15:45–16:10

Extensive lexical knowledge is necessary for temporal analysis and planning tasks. We ad- dress in this paper a lexical setting that allows for the straightforward incorporation of rich features and structural constraints. We explore a lexical event ordering task, namely determining the likely temporal order of events based solely on the identity of their predicates and arguments. We propose an "edge-factored" model for the task that decomposes over the edges of the event graph. We learn it using the structured perceptron. As lexical tasks require large amounts of text, we do not attempt manual annotation and instead use the textual order of events in a domain where this order is aligned with their temporal order, namely cooking recipes.

[TACL] Entity disambiguation with web links
Andrew Chisholm and Ben Hachey 16:10–16:35

Entity disambiguation with Wikipedia relies on structured information from redirect pages, article text, inter-article links, and categories. We explore whether web links can replace a curated encyclopaedia, obtaining entity prior, name, context, and coherence models from a corpus of web pages with links to Wikipedia. Experiments compare web link models to Wikipedia models on well-known CoNLL and TAC data sets. Results show that using 34 million web links approaches Wikipedia performance. Combining web link and Wikipedia models produces the best-known disambiguation accuracy of 88.7 on standard newswire test data.

[TACL] A Joint Model for Entity Analysis: Coreference, Typing, and Linking
Greg Durrett and Dan Klein 16:35–17:00

We present a joint model of three core tasks in the entity analysis stack: coreference resolution (within-document clustering), named entity recognition (coarse semantic typing), and entity linking (matching to Wikipedia entities). Our model is formally a structured conditional random field. Unary factors encode local features from strong baselines for each task. We then add binary and ternary factors to capture cross-task interactions, such as the constraint that coreferent mentions have the same semantic type. On the ACE 2005 and OntoNotes datasets, we achieve state-of-the-art results for all three tasks. Moreover, joint modeling improves performance on each task over strong independent baselines.

Session 7C: Machine Translation (Long Papers)

Plaza Ballroom F *Chair: Hany Hassan*

Bag-of-Words Forced Decoding for Cross-Lingual Information Retrieval
Felix Hieber and Stefan Riezler 15:45–16:10

Current approaches to cross-lingual information retrieval (CLIR) rely on standard retrieval models into which query translations by statistical machine translation (SMT) are integrated at varying degree. In this paper, we present an attempt to turn this situation on its head: Instead of the retrieval aspect, we emphasize the translation component in CLIR. We perform search by using an SMT decoder in forced decoding mode to produce a bag-of-words representation of the target documents to be ranked. The SMT model is extended by retrieval-specific features that are optimized jointly with standard translation features for a ranking objective. We find significant gains over the state-of-the-art in a large-scale evaluation on cross-lingual search in the domains patents and Wikipedia.

Accurate Evaluation of Segment-level Machine Translation Metrics
Yvette Graham, Timothy Baldwin, and Nitika Mathur 16:10–16:35

Evaluation of segment-level machine translation metrics is currently hampered by: (1) low inter-annotator agreement levels in human assessments; (2) lack of an effective mechanism for evaluation of translations of equal quality; and (3) lack of methods of significance testing improvements over a baseline. In this paper, we provide solutions to each of these challenges and outline a new human evaluation methodology aimed specifically at assessment of segment-level metrics. We replicate the human evaluation component of WMT-13 and reveal that the current state-of-the-art performance of segment-level metrics is better than previously believed. Three segment-level metrics — Meteor, nLepor and sentBLEU-moses — are found to correlate with human assessment at a level not significantly outperformed by any other metric in both the individual language pair assessment for Spanish to English and the aggregated set of 9 language pairs.

Leveraging Small Multilingual Corpora for SMT Using Many Pivot Languages
Raj Dabre, Fabien Cromieres, Sadao Kurohashi, and Pushpak Bhattacharyya 16:35–17:00

We present our work on leveraging multilingual parallel corpora of small sizes for Statistical Machine Translation between Japanese and Hindi using multiple pivot languages. In our setting, the source and target part of the corpus remains the same, but we show that using several different pivot to extract phrase pairs from these source and target parts lead to large BLEU improvements. We focus on a variety of ways to exploit phrase tables generated using multiple pivots to support a direct source-target phrase table. Our main method uses the Multiple Decoding Paths (MDP) feature of Moses, which we empirically verify as the best compared to the other methods we used. We compare and contrast our various results to show that one can overcome the limitations of small corpora by using as many pivot languages as possible in a multilingual setting. Most importantly, we show that such pivoting aids in learning of additional phrase pairs which are not learned when the direct source-target corpus is small. We obtained improvements of up to 3 BLEU points using multiple pivots for Japanese to Hindi translation compared to when only one pivot is used. To the best of our knowledge, this work is also the first of its kind to attempt the simultaneous utilization of 7 pivot languages at decoding time.

Demo Session A

Time: *5:00–6:30* Location: Plaza Ballroom A, B, & C

Two Practical Rhetorical Structure Theory Parsers
Mihai Surdeanu, Tom Hicks, and Marco Antonio Valenzuela-Escarcega

We describe the design, development, and API for two discourse parsers for Rhetorical Structure Theory. The two parsers use the same underlying framework, but one uses features that rely on dependency syntax, produced by a fast shift-reduce parser, whereas the other uses a richer feature space, including both constituent- and dependency-syntax and coreference information, produced by the Stanford CoreNLP toolkit. Both parsers obtain state-of-the-art performance, and use a very simple API consisting of, minimally, two lines of Scala code. We accompany this code with a visualization library that runs the two parsers in parallel, and displays the two generated discourse trees side by side, which provides an intuitive way of comparing the two parsers.

Analyzing and Visualizing Coreference Resolution Errors
Sebastian Martschat, Thierry Göckel, and Michael Strube

We present a toolkit for coreference resolution error analysis. It implements a recently proposed analysis framework and contains rich components for analyzing and visualizing recall and precision errors.

hyp: A Toolkit for Representing, Manipulating, and Optimizing Hypergraphs
Markus Dreyer and Jonathan Graehl

We present hyp, an open-source toolkit for the representation, manipulation, and optimization of weighted directed hypergraphs. hyp provides compose, project, invert functionality, k-best path algorithms, the inside and outside algorithms, and more. Finite-state machines are modeled as a special case of directed hypergraphs. hyp consists of a C++ API, as well as a command line tool, and is available for download at github.com/sdl-research/hyp.

Enhancing Instructor-Student and Student-Student Interactions with Mobile Interfaces and Summarization
Wencan Luo, Xiangmin Fan, Muhsin Menekse, Jingtao Wang, and Diane Litman

Educational research has demonstrated that asking students to respond to reflection prompts can increase interaction between instructors and students, which in turn can improve both teaching and learning especially in large classrooms. However, administering an instructor's prompts, collecting the students' responses, and summarizing these responses for both instructors and students is challenging and expensive. To address these challenges, we have developed an application called CourseMIRROR (Mobile In-situ Reflections and Review with Optimized Rubrics). CourseMIRROR uses a mobile interface to administer prompts and collect reflective responses for a set of instructor-assigned course lectures. After collection, CourseMIRROR automatically summarizes the reflections with an extractive phrase summarization method, using a clustering algorithm to rank extracted phrases by student coverage. Finally, CourseMIRROR presents the phrase summary to both instructors and students to help them understand the difficulties and misunderstandings encountered.

RExtractor: a Robust Information Extractor
Vincent Kríž and Barbora Hladka

The RExtractor system is an information extractor that processes input documents by natural language processing tools and consequently queries the parsed sentences to extract a knowledge base of entities and their relations. The extraction queries are designed manually using a tool that enables natural graphical representation of queries over dependency trees. A workflow of the system is designed to be language and domain independent. We demonstrate RExtractor on Czech and English legal documents.

An AMR parser for English, French, German, Spanish and Japanese and a new AMR-annotated corpus
Lucy Vanderwende, Arul Menezes, and Chris Quirk

In this demonstration, we will present our online parser that allows users to submit any sentence and obtain an analysis following the specification of AMR (Banarescu et al., 2014) to a large extent. This AMR analysis is generated by a small set of rules that convert a native Logical Form analysis provided by a pre-existing parser into the AMR format. While we demonstrate the performance of our AMR parser on data sets annotated by the LDC, we will focus attention in the demo on the following two areas: 1) we will make available AMR

annotations for the data sets that were used to develop our parser, to serve as a supplement to the LDC data sets, and 2) we will demonstrate AMR parsers for German, French, Spanish and Japanese that make use of the same small set of LF-to-AMR conversion rules.

ICE: Rapid Information Extraction Customization for NLP Novices
Yifan He and Ralph Grishman

We showcase ICE, an Integrated Customization Environment for Information Extraction. ICE is an easy tool for non-NLP experts to rapidly build customized IE systems for a new domain.

AMRICA: an AMR Inspector for Cross-language Alignments
Naomi Saphra and Adam Lopez

Abstract Meaning Representation (AMR), an annotation scheme for natural language semantics, has drawn attention for its simplicity and representational power. Because AMR annotations are not designed for human readability, we present AMRICA, a visual aid for exploration of AMR annotations. AMRICA can visualize an AMR or the difference between two AMRs to help users diagnose interannotator disagreement or errors from an AMR parser. AMRICA can also automatically align and visualize the AMRs of a sentence and its translation in a parallel text. We believe AMRICA will simplify and streamline exploratory research on cross-lingual AMR corpora.

Ckylark: A More Robust PCFG-LA Parser
Yusuke Oda, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura

This paper describes Ckylark, a PCFG-LA style phrase structure parser that is more robust than other parsers in the genre. PCFG-LA parsers are known to achieve highly competitive performance, but sometimes the parsing process fails completely, and no parses can be generated. Ckylark introduces three new techniques that prevent possible causes for parsing failure: outputting intermediate results when coarse-to-fine analysis fails, smoothing lexicon probabilities, and scaling probabilities to avoid underflow. An experiment shows that this allows millions of sentences can be parsed without any failures, in contrast to other publicly available PCFG-LA parsers. Ckylark is implemented in C++, and is available open-source under the LGPL license.

ELCO3: Entity Linking with Corpus Coherence Combining Open Source Annotators
Pablo Ruiz, Thierry Poibeau, and Fréderique Mélanie

Entity Linking (EL) systems' performance is uneven across corpora or depending on entity types. To help overcome this issue, we propose an EL workflow that combines the outputs of several open source EL systems, and selects annotations via weighted voting. The results are displayed on a UI that allows the users to navigate the corpus and to evaluate annotation quality based on several metrics.

SETS: Scalable and Efficient Tree Search in Dependency Graphs
Juhani Luotolahti, Jenna Kanerva, Sampo Pyysalo, and Filip Ginter

We present a syntactic analysis query toolkit geared specifically towards massive dependency parsebanks and morphologically rich languages. The query language allows arbitrary tree queries, including negated branches, and is suitable for querying analyses with rich morphological annotation. Treebanks of over a million words can be comfortably queried on a low-end netbook, and a parsebank with over 100M words on a single consumer-grade server. We also introduce a web-based interface for interactive querying. All contributions are available under open licenses.

Visualizing Deep-Syntactic Parser Output
Juan Soler-Company, Miguel Ballesteros, Bernd Bohnet, Simon Mille, and Leo Wanner

"Deep-syntactic" dependency structures bridge the gap between the surface-syntactic structures as produced by state-of-the-art dependency parsers and semantic logical forms in that they abstract away from surface-syntactic idiosyncrasies, but still keep the linguistic structure of a sentence. They have thus a great potential for such downstream applications as machine translation and summarization. In this demo paper, we propose an online version of a deep-syntactic parser that outputs deep-syntactic structures from plain sentences and visualizes them using the Brat tool. Along with the deep-syntactic structures, the user can also inspect the visual presentation of the surface-syntactic structures that serve as input to the deep-syntactic parser and that are produced by the joint tagger and syntactic transition-based parser ran in the pipeline before the deep-syntactic parser.

WOLFE: An NLP-friendly Declarative Machine Learning Stack
Sameer Singh, Tim Rocktäschel, Luke Hewitt, Jason Naradowsky, and Sebastian Riedel

Developing machine learning algorithms for natural language processing (NLP) applications is inherently an iterative process, involving a continuous refinement of the choice of model, engineering of features, selection of inference algorithms, search for the right hyper- parameters, and error analysis. Existing probabilistic program languages (PPLs) only provide partial solutions; most of them do not support commonly used models such as matrix factorization or neural networks, and do not facilitate interactive and iterative programming that is crucial for rapid development of these models. In this demo we introduce WOLFE, a stack designed to facilitate the development of NLP applications: (1) the WOLFE language allows the user to concisely define complex models, enabling easy modification and extension, (2) the WOLFE interpreter transforms declarative machine learning code into automatically differentiable terms or, where applicable, into factor graphs that allow for complex models to be applied to real-world applications, and (3) the WOLFE IDE provides a number of different visual and interactive elements, allowing intuitive exploration and editing of the data representations, the underlying graphical models, and the execution of the inference algorithms.

Demo Session B

Time: *6:30–8:00* Location: Plaza Ballroom A, B, & C

Lean Question Answering over Freebase from Scratch
Xuchen Yao

For the task of question answering (QA) over Freebase on the WebQuestions dataset (Berant et al., 2013), we found that 85% of all questions (in the training set) can be directly answered via a single binary relation. Thus we turned this task into slot-filling for <question topic, relation, answer> tuples: predicting relations to get answers given a question's topic. We design efficient data structures to identify question topics organically from 46 million Freebase topic names, without employing any NLP processing tools. Then we present a lean QA system that runs in real time (in offline batch testing it answered two thousand questions in 51 seconds on a laptop). The system also achieved 7.8% better F1 score (harmonic mean of average precision and recall) than the previous state of the art.

A Web Application for Automated Dialect Analysis
Sravana Reddy and James Stanford

Sociolinguists are regularly faced with the task of measuring phonetic features from speech, which involves manually transcribing audio recordings – a major bottleneck to analyzing large collections of data. We harness automatic speech recognition to build an online end-to-end web application where users upload untranscribed speech collections and receive formant measurements of the vowels in their data. We demonstrate this tool by using it to automatically analyze President Barack Obama's vowel pronunciations.

An Open-source Framework for Multi-level Semantic Similarity Measurement
Mohammad Taher Pilehvar and Roberto Navigli

We present an open source, freely available Java implementation of Align, Disambiguate, and Walk (ADW), a state-of-the-art approach for measuring semantic similarity based on the Personalized PageRank algorithm. A pair of linguistic items, such as phrases or sentences, are first disambiguated using an alignment-based disambiguation technique and then modeled using random walks on the WordNet graph. ADW provides three main advantages: (1) it is applicable to all types of linguistic items, from word senses to texts; (2) it is all-in-one, i.e., it does not need any additional resource, training or tuning; and (3) it has proven to be highly reliable at different lexical levels and multiple evaluation benchmarks. We are releasing the source code at https://github.com/pilehvar/adw/. We also provide at http://lcl.uniroma1.it/adw/ a Web interface and a Java API that can be seamlessly integrated into other NLP systems requiring semantic similarity measurement.

Brahmi-Net: A transliteration and script conversion system for languages of the Indian subcontinent
Anoop Kunchukuttan, Ratish Puduppully, and Pushpak Bhattacharyya

We present Brahmi-Net - an online system for transliteration and script conversion for all major Indian language pairs (306 pairs). The system covers 13 Indo-Aryan languages, 4 Dravidian languages and English. For training the transliteration systems, we mined parallel transliteration corpora from parallel translation cor-

pora using an unsupervised method and trained statistical transliteration systems using the mined corpora. Languages which do not have a parallel corpora are supported by transliteration through a bridge language. Our script conversion system supports conversion between all Brahmi-derived scripts as well as ITRANS romanization scheme. For this, we leverage co-ordinated Unicode ranges between Indic scripts and use an extended ITRANS encoding for transliterating between English and Indic scripts. The system also provides top-k transliterations and simultaneous transliteration into multiple output languages. We provide a Python as well as REST API to access these services. The API and the mined transliteration corpus are made available for research use under an open source license.

A Concrete Chinese NLP Pipeline
Nanyun Peng, Francis Ferraro, Mo Yu, Nicholas Andrews, Jay DeYoung, Max Thomas, Matthew R. Gormley, Travis Wolfe, Craig Harman, Benjamin Van Durme, and Mark Dredze

Natural language processing research increasingly relies on the output of a variety of syntactic and semantic analytics. Yet integrating output from multiple analytics into a single framework can be time consuming and slow research progress. We present a Chinese Concrete NLP Pipeline: an NLP stack built using a series of open source systems integrated based on the Concrete data schema. Our pipeline includes data ingest, word segmentation, part of speech tagging, parsing, named entity recognition, relation extraction and cross document coreference resolution. Additionally, we integrate a tool for visualizing these annotations as well as allowing for the manual annotation of new data. We release our pipeline to the research community to facilitate work on Chinese language tasks that require rich linguistic annotations.

CroVeWA: Crosslingual Vector-Based Writing Assistance
Hubert Soyer, Goran Topić, Pontus Stenetorp, and Akiko Aizawa

We present an interactive web-based writing assistance system that is based on recent advances in crosslingual compositional distributed semantics. Given queries in Japanese or English, our system can retrieve semantically related sentences from high quality English corpora. By employing crosslingually constrained vector space models to represent phrases, our system naturally sidesteps several difficulties that would arise from direct word-to-text matching, and is able to provide novel functionality like the visualization of semantic relationships between phrases interlingually and intralingually.

Online Readability and Text Complexity Analysis with TextEvaluator
Diane Napolitano, Kathleen Sheehan, and Robert Mundkowsky

We have developed the TextEvaluator system for providing text complexity and Common Core-aligned readability information. Detailed text complexity information is provided by eight component scores, presented in such a way as to aid in the user's understanding of the overall readability metric, which is provided as a holistic score on a scale of 100 to 2000. The user may select a targeted US grade level and receive additional analysis relative to it. This and other capabilities are accessible via a feature-rich front-end, located at http://texteval-pilot.ets.org/TextEvaluator/.

Natural Language Question Answering and Analytics for Diverse and Interlinked Datasets
Dezhao Song, Frank Schilder, Charese Smiley, and Chris Brew

Previous systems for natural language questions over complex linked datasets require the user to enter a complete and well-formed question, and present the answers as raw lists of entities. Using a feature-based grammar with a full formal semantics, we have developed a system that is able to support rich autosuggest, and to deliver dynamically generated analytics for each result that it returns.

WriteAhead2: Mining Lexical Grammar Patterns for Assisted Writing
Jim Chang and Jason Chang

This paper describes WriteAhead2, an interactive writing environment that provides lexical and grammatical suggestions for second language learners, and helps them write fluently and avoid common writing errors. The method involves learning phrase templates from dictionary examples, and extracting grammar patterns with example phrases from an academic corpus. At run-time, as the user types word after word, the actions trigger a list after list of suggestions. Each successive list contains grammar patterns and examples, most relevant to the half-baked sentence. WriteAhead2 facilitates steady, timely, and spot-on interactions between learner writers and relevant information for effective assisted writing. Preliminary experiments show that WriteAhead2 has the potential to induce better writing and improve writing skills.

Question Answering System using Multiple Information Source and Open Type Answer Merge
Seonyeong Park, Soonchoul Kwon, Byungsoo Kim, Sangdo Han, Hyosup Shim, and Gary Geunbae Lee

This paper presents a multi-strategy and multi-source question answering (QA) system that can use multiple strategies to both answer natural language (NL) questions and respond to keywords. We use multiple information sources including curated knowledge base, raw text, auto-generated triples, and NL processing results. We develop open semantic answer type detector for answer merging and improve previous developed single QA modules such as knowledge base based QA, information retrieval based QA.

Using Word Semantics To Assist English as a Second Language Learners
Mahmoud Azab, Chris Hokamp, and Rada Mihalcea

We introduce an interactive interface that aims to help English as a Second Language (ESL) students overcome language related hindrances while reading a text. The interface allows the user to find supplementary information on selected difficult words. The interface is empowered by our lexical substitution engine that provides context-based synonyms for difficult words. We also provide a practical solution for a real-world usage scenario. We demonstrate using the lexical substitution engine – as a browser extension that can annotate and disambiguate difficult words on any webpage.

Social Event

Tuesday, June 2, 2015, 8pm–11pm

Grand Ballroom

The NAACL 2015 Social and Networking Event will be held immediately following the Tuesday Poster Session and dinner in the Grand Ballroom at the hotel. Here you will enjoy desserts, coffee and tea, and a cash bar. Bring your boots and hats and follow along with our dance instructors to learn the latest country line dancing, then practice your moves to the sounds of a local DJ (who will play other sounds if you tire of the country twang). Enjoy networking with colleagues and have a relaxing evening!

(We are trying a second-time experiment in holding a second poster dinner and open-to-all desserts, drinks, and dancing social event combination in place of the usual Banquet).

We hope to make your conference experience not only enlightening but also entertaining and enjoyable!

5

Main Conference: Wednesday, June 3

Overview

7:30 – 9:00	**Registration and Breakfast**		*Plaza Exhibit All*
9:00 – 10:10	**Invited Talk: "A Quest for Visual Intelligence in Computers" (Fei-fei Li)** *Plaza Ballroom A, B, & C*		
10:10 – 10:40	**Break**		*Plaza Exhibit All*
	Session 8		
10:40 – 11:55	NLP for Web, Social Media and Social Sciences (Long + TACL Papers) *Plaza Ballroom A & B*	Language and Vision (Long + TACL Papers) *Plaza Ballroom D & E*	Machine Translation (Long + TACL Papers) *Plaza Ballroom F*
11:55 – 1:00	**Lunch**		
1:00 – 2:00	**Business Meeting**		*Plaza Ballroom A & B*
	Session 9		
2:00 – 3:15	Lexical Semantics and Sentiment Analysis (Long Papers) *Plaza Ballroom A & B*	NLP-enabled Technology (Long + TACL Papers) *Plaza Ballroom D & E*	Linguistic and Psycholinguistic Aspects of CL (Long Papers) *Plaza Ballroom F*
3:15 – 3:45	**Break**		*Plaza Exhibit All*
3:45 – 5:15	**Best Paper Plenary Session**		*Plaza Ballroom A, B, & C*
5:15 – 5:30	**Best Paper Awards and Closing Remarks**		*Plaza Ballroom A, B, & C*

Keynote Address: Fei-Fei Li

"A Quest for Visual Intelligence in Computers"

Wednesday, June 3, 2015, 9:00–10:10am

Plaza Ballroom A, B, & C

Abstract: More than half of the human brain is involved in visual processing. While it took mother nature billions of years to evolve and deliver us a remarkable human visual system, computer vision is one of the youngest disciplines of AI, born with the goal of achieving one of the loftiest dreams of AI. The central problem of computer vision is to turn millions of pixels of a single image into interpretable and actionable concepts so that computers can understand pictures just as well as humans do, from objects, to scenes, activities, events and beyond. Such technology will have a fundamental impact in almost every aspect of our daily life and the society as a whole, ranging from e-commerce, image search and indexing, assistive technology, autonomous driving, digital health and medicine, surveillance, national security, robotics and beyond. In this talk, I will give an overview of what computer vision technology is about and its brief history. I will then discuss some of the recent work from my lab towards large scale object recognition and visual scene story telling. I will particularly emphasize on what we call the "three pillars" of AI in our quest for visual intelligence: data, learning and knowledge. Each of them is critical towards the final solution, yet dependent on the other. This talk draws upon a number of projects ongoing at the Stanford Vision Lab.

Biography: Dr. Fei-Fei Li is an Associate Professor in the Computer Science Department at Stanford, and the Director of the Stanford Artificial Intelligence Lab and the Stanford Vision Lab. Her research areas are in machine learning, computer vision and cognitive and computational neuroscience, with an emphasis on Big Data analysis. Dr. Fei-Fei Li has published more than 100 scientific articles in top-tier journals and conferences, including Nature, PNAS, Journal of Neuroscience, CVPR, ICCV, NIPS, ECCV, IJCV, IEEE-PAMI, etc. Dr. Fei-Fei Li obtained her B.A. degree in physics from Princeton in 1999 with High Honors, and her PhD degree in electrical engineering from California Institute of Technology (Caltech) in 2005. She joined Stanford in 2009 as an assistant professor, and was promoted to associate professor with tenure in 2012. Prior to that, she was on faculty at Princeton University (2007-2009) and University of Illinois Urbana-Champaign (2005-2006). Dr. Fei-Fei Li is a speaker at TED2015 main conference, a recipient of the 2014 IBM Faculty Fellow Award, 2011 Alfred Sloan Faculty Award, 2012 Yahoo Labs FREP award, 2009 NSF CAREER award, the 2006 Microsoft Research New Faculty Fellowship and a number of Google Research awards. Work from Fei-Fei's lab have been featured in a number of popular press magazines and newspapers including New York Times, Wired Magazine, and New Scientists.

Session 8 Overview – Wednesday, June 3, 2015

Track A	Track B	Track C	
NLP for Web, Social Media and Social Sciences (Long + TACL Papers) Plaza Ballroom A & B	*Language and Vision (Long + TACL Papers)* Plaza Ballroom D & E	*Machine Translation (Long + TACL Papers)* Plaza Ballroom F	
Testing and Comparing Computational Approaches for Identifying the Language of Framing in Political News *Baumer, Elovic, Qin, Polletta, and Gay*	Translating Videos to Natural Language Using Deep Recurrent Neural Networks *Venugopalan, Xu, Donahue, Rohrbach, Mooney, and Saenko*	A Comparison of Update Strategies for Large-Scale Maximum Expected BLEU Training *Wuebker, Muehr, Lehnen, Peitz, and Ney*	10:40
[TACL] Extracting Lexically Divergent Paraphrases from Twitter *Xu, Ritter, Callison-Burch, Dolan, and Ji*	[TACL] A Bayesian Model of Grounded Color Semantics *McMahan and Stone*	[TACL] Gappy Pattern Matching on GPUs for On-Demand Extraction of Hierarchical Translation Grammars *He, Lin, and Lopez*	11:05
Echoes of Persuasion: The Effect of Euphony in Persuasive Communication *Guerini, Özbal, and Strapparava*	Learning to Interpret and Describe Abstract Scenes *Gilberto Mateos Ortiz, Wolff, and Lapata*	Learning Translation Models from Monolingual Continuous Representations *Zhao, Hassan, and Auli*	11:30

Parallel Session 8

Session 8A: NLP for Web, Social Media and Social Sciences (Long + TACL Papers)

Plaza Ballroom A & B *Chair: Philip Resnik*

Testing and Comparing Computational Approaches for Identifying the Language of Framing in Political News

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay 10:40–11:05

The subconscious influence of framing on perceptions of political issues is well-document in political science and communication research. A related line of work suggests that drawing attention to framing may help reduce such framing effects by enabling frame reflection, critical examination of the framing underlying an issue. However, definite guidance on how to identify framing does not exist. This paper presents a technique for identifying frame-invoking language. The paper first describes a human subjects pilot study that explores how individuals identify framing and informs the design of our technique. The paper then describes our data collection and annotation approach. Results show that the best performing classifiers achieve performance comparable to that of human annotators, and they indicate which aspects of language most pertain to framing. Both technical and theoretical implications are discussed.

[TACL] Extracting Lexically Divergent Paraphrases from Twitter

Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan, and Yangfeng Ji 11:05–11:30

We present MultiP (Multi-instance Learning Paraphrase Model), a new model suited to identify paraphrases within the short messages on Twitter. We jointly model paraphrase relations between word and sentence pairs and assume only sentence-level annotations during learning. Using this principled latent variable model alone, we achieve the performance competitive with a state-of-the-art method which combines a latent space model with a feature-based supervised classifier. Our model also captures lexically divergent paraphrases that differ from yet complement previous methods; combining our model with previous work significantly outperforms the state-of-the-art. In addition, we present a novel annotation methodology that has allowed us to crowdsource a paraphrase corpus from Twitter. We make this new dataset available to the research community.

Echoes of Persuasion: The Effect of Euphony in Persuasive Communication

Marco Guerini, Gözde Özbal, and Carlo Strapparava 11:30–11:55

While the effect of various lexical, syntactic, semantic and stylistic features have been addressed in persuasive language from a computational point of view, the persuasive effect of phonetics has received little attention. By modeling a notion of euphony and analyzing four datasets comprising persuasive and non-persuasive sentences in different domains (political speeches, movie quotes, slogans and tweets), we explore the impact of sounds on different forms of persuasiveness. We conduct a series of analyses and prediction experiments within and across datasets. Our results highlight the positive role of phonetic devices on persuasion.

Session 8B: Language and Vision (Long + TACL Papers)

Plaza Ballroom D & E *Chair: Luke Zettlemoyer*

Translating Videos to Natural Language Using Deep Recurrent Neural Networks
Subhashini Venugopalan, Huijuan Xu, Jeff Donahue, Marcus Rohrbach, Raymond Mooney, and Kate Saenko 10:40–11:05

Solving the visual symbol grounding problem has long been a goal of artificial intelligence. The field appears to be advancing closer to this goal with recent breakthroughs in deep learning for natural language grounding in static images. In this paper, we propose to translate videos directly to sentences using a unified deep neural network with both convolutional and recurrent structure. Described video datasets are scarce, and most existing methods have been applied to toy domains with a small vocabulary of possible words. By transferring knowledge from 1.2M+ images with category labels and 100,000+ images with captions, our method is able to create sentence descriptions of open-domain videos with large vocabularies. We compare our approach with recent work using language generation metrics, subject, verb, and object prediction accuracy, and a human evaluation.

[TACL] A Bayesian Model of Grounded Color Semantics
Brian McMahan and Matthew Stone 11:05–11:30

Natural language meanings allow speakers to encode important real-world distinctions, but corpora of grounded language use also reveal that speakers categorize the world in different ways and describe situations with different terminology. To learn meanings from data, we therefore need to link underlying representations of meaning to models of speaker judgment and speaker choice. This paper describes a new approach to this problem: we model variability through uncertainty in categorization boundaries and distributions over preferred vocabulary. We apply the approach to a large data set of color descriptions, where statistical evaluation documents its accuracy. The results are available as a Lexicon of Uncertain Color Standards (LUX), which supports future efforts in grounded language understanding and generation by probabilistically mapping 829 English color descriptions to potentially context-sensitive regions in HSV color space.

Learning to Interpret and Describe Abstract Scenes
Luis Gilberto Mateos Ortiz, Clemens Wolff, and Mirella Lapata 11:30–11:55

Given a (static) scene, a human can effortlessly describe what is going on (who is doing what to whom, how, and why). The process requires knowledge about the world, how it is perceived, and described. In this paper we study the problem of interpreting and verbalizing visual information using abstract scenes created from collections of clip art images. We propose a model inspired by machine translation operating over a large parallel corpus of visual relations and linguistic descriptions. We demonstrate that this approach produces human-like scene descriptions which are both fluent and relevant, outperforming a number of competitive alternatives based on templates and sentence-based retrieval.

Session 8C: Machine Translation (Long + TACL Papers)

Plaza Ballroom F *Chair: Haitao Mi*

A Comparison of Update Strategies for Large-Scale Maximum Expected BLEU Training
Joern Wuebker, Sebastian Muehr, Patrick Lehnen, Stephan Peitz, and Hermann Ney 10:40–11:05

This work presents a flexible and efficient discriminative training approach for statistical machine translation. We propose to use the RPROP algorithm for optimizing a maximum expected BLEU objective and experimentally compare it to several other updating schemes. It proves to be more efficient and effective than the previously proposed growth transformation technique and also yields better results than stochastic gradient descent and AdaGrad. We also report strong empirical results on two large scale tasks, namely BOLT Chinese->English and WMT German->English, where our final systems outperform results reported by Setiawan and Zhou (2013) and on matrix.statmt.org. On the WMT task, discriminative training is performed on the full training data of 4M sentence pairs, which is unsurpassed in the literature.

[TACL] Gappy Pattern Matching on GPUs for On-Demand Extraction of Hierarchical Translation Grammars
Hua He, Jimmy Lin, and Adam Lopez 11:05–11:30

Grammars for machine translation can be materialized on demand by finding source phrases in an indexed parallel corpus and extracting their translations. This approach is limited in practical applications by the computational expense of online lookup and extraction. For phrase-based models, recent work has shown that on-demand grammar extraction can be greatly accelerated by parallelization on general purpose graphics processing units (GPUs), but these algorithms do not work for hierarchical models, which require matching patterns that contain gaps. We address this limitation by presenting a novel GPU algorithm for on-demand hierarchical grammar extraction that is at least an order of magnitude faster than a comparable CPU algorithm when processing large batches of sentences. In terms of end-to-end translation, with decoding on the CPU, we increase throughput by roughly two thirds on a standard MT evaluation dataset. The GPU necessary to achieve these improvements increases the cost of a server by about a third. We believe that GPU-based extraction of hierarchical grammars is an attractive proposition, particularly for MT applications that demand high throughput.

Learning Translation Models from Monolingual Continuous Representations
Kai Zhao, Hany Hassan, and Michael Auli 11:30–11:55

Translation models often fail to generate good translations for infrequent words or phrases. Previous work attacked this problem by inducing new translation rules from monolingual data with a semi-supervised algorithm. However, this approach does not scale very well since it is very computationally expensive to generate new translation rules for only a few thousand sentences. We propose a much faster and simpler method that directly hallucinates translation rules for infrequent phrases based on phrases with similar continuous representations for which a translation is known. To speed up the retrieval of similar phrases, we investigate approximated nearest neighbor search with redundant bit vectors which we find to be three times faster and significantly more accurate than locality sensitive hashing. Our approach of learning new translation rules improves a phrase-based baseline by up to 1.6 BLEU on Arabic-English translation, it is three-orders of magnitudes faster than existing semi-supervised methods and 0.5 BLEU more accurate.

NAACL Business Meeting

Date: Wednesday, June 3, 2015
Time: 1:00–2:00 PM
Venue: Plaza Ballroom A & B

Chair: Hal Daumé III

All attendees are encouraged to participate in the business meeting.

Session 9 Overview – Wednesday, June 3, 2015

	Track A	Track B	Track C
	Lexical Semantics and Sentiment Analysis (Long Papers)	*NLP-enabled Technology (Long + TACL Papers)*	*Linguistic and Psycholinguistic Aspects of CL (Long Papers)*
	Plaza Ballroom A & B	Plaza Ballroom D & E	Plaza Ballroom F
2:00	A Corpus and Model Integrating Multiword Expressions and Supersenses *Schneider and Smith*	How to Memorize a Random 60-Bit String *Ghazvininejad and Knight*	A Bayesian Model for Joint Learning of Categories and their Features *Frermann and Lapata*
2:25	Good News or Bad News: Using Affect Control Theory to Analyze Readers' Reaction Towards News Articles *Alhothali and Hoey*	[TACL] Building a State-of-the-Art Grammatical Error Correction System *Rozovskaya and Roth*	Shared common ground influences information density in microblog texts *Doyle and Frank*
2:50	Do We Really Need Lexical Information? Towards a Top-down Approach to Sentiment Analysis of Product Reviews *Otmakhova and Shin*	[TACL] Predicting the Difficulty of Language Proficiency Tests *Beinborn, Zesch, and Gurevych*	Hierarchic syntax improves reading time prediction *Schijndel and Schuler*

Parallel Session 9

Session 9A: Lexical Semantics and Sentiment Analysis (Long Papers)

Plaza Ballroom A & B *Chair: Saif Mohammed*

A Corpus and Model Integrating Multiword Expressions and Supersenses
Nathan Schneider and Noah A. Smith 14:00–14:25

This paper introduces a task of identifying and semantically classifying lexical expressions in context. We investigate the online reviews genre, adding semantic supersense annotations to a 55,000 word English corpus that was previously annotated for multiword expressions. The noun and verb supersenses apply to full lexical expressions, whether single- or multiword. We then present a sequence tagging model that jointly infers lexical expressions and their supersenses. Results show that even with our relatively small training corpus in a noisy domain, the joint task can be performed to attain 70% class labeling F1.

Good News or Bad News: Using Affect Control Theory to Analyze Readers' Reaction Towards News Articles
Areej Alhothali and Jesse Hoey 14:25–14:50

This paper proposes a novel approach to sentiment analysis that leverages work in sociology on symbolic interactionism. The proposed approach uses Affect Control Theory (ACT) to analyze readers' sentiment towards factual (objective) content and towards its entities (subject and object). ACT is a theory of affective reasoning that uses empirically derived equations to predict the sentiments and emotions that arise from events. This theory relies on several large lexicons of words with affective ratings in a three-dimensional space of evaluation, potency, and activity (EPA). The equations and lexicons of ACT were evaluated on a newly collected news-headlines corpus. ACT lexicon was expanded using a label propagation algorithm, resulting in 86,604 new words. The predicted emotions for each news headline was then computed using the augmented lexicon and ACT equations. The results had a precision of 82%, 79%, and 68% towards the event, the subject, and object, respectively. These results are significantly higher than those of standard sentiment analysis techniques.

Do We Really Need Lexical Information? Towards a Top-down Approach to Sentiment Analysis of Product Reviews
Yulia Otmakhova and Hyopil Shin 14:50–15:15

Most of the current approaches to sentiment analysis of product reviews are dependent on lexical sentiment information and proceed in a bottom-up way, adding new layers of features to lexical data. In this paper, we maintain that a typical product review is not a bag of sentiments, but a narrative with an underlying structure and reoccurring patterns, which allows us to predict its sentiments knowing only its general polarity and discourse cues that occur in it. We hypothesize that knowing only the review's score and its discourse patterns would allow us to accurately predict the sentiments of its individual sentences. The experiments we conducted prove this hypothesis and show a substantial improvement over the lexical baseline.

Session 9B: NLP-enabled Technology (Long + TACL Papers)

Plaza Ballroom D & E *Chair: Brendan O'Connor*

How to Memorize a Random 60-Bit String
Marjan Ghazvininejad and Kevin Knight 14:00–14:25

User-generated passwords tend to be memorable, but not secure. A random, computer-generated 60-bit string is much more secure. However, users cannot memorize random 60-bit strings. In this paper, we investigate methods for converting arbitrary bit strings into English word sequences (both prose and poetry), and we study their memorability and other properties.

[TACL] Building a State-of-the-Art Grammatical Error Correction System
Alla Rozovskaya and Dan Roth 14:25–14:50

This paper identifies and examines the key principles underlying building a state-of-the-art grammatical error correction system. We do this by analyzing the Illinois system that placed first among seventeen teams in the recent CoNLL-2013 shared task on grammatical error correction. The system focuses on five different types of errors common among non-native English writers. We describe four design principles that are relevant for correcting all of these errors, analyze the system along these dimensions, and show how each of these dimensions contributes to the performance.

[TACL] Predicting the Difficulty of Language Proficiency Tests
Lisa Beinborn, Torsten Zesch, and Iryna Gurevych 14:50–15:15

Language proficiency tests are used to evaluate and compare the progress of language learners. We present an approach for automatic difficulty prediction of C-tests that performs on par with human experts. On the basis of detailed analysis of newly collected data, we develop a model for C-test difficulty introducing four dimensions: solution difficulty, candidate ambiguity, inter-gap dependency, and paragraph difficulty. We show that cues from all four dimensions contribute to C-test difficulty.

Session 9C: Linguistic and Psycholinguistic Aspects of CL (Long Papers)

Plaza Ballroom F *Chair: William Schuler*

A Bayesian Model for Joint Learning of Categories and their Features

Lea Frermann and Mirella Lapata 14:00–14:25

Categories such as ANIMAL or FURNITURE are acquired at an early age and play an important role in processing, organizing, and conveying world knowledge. Theories of categorization largely agree that categories are characterized by features such as function or appearance and that feature and category acquisition go hand-in-hand, however previous work has considered these problems in isolation. We present the first model that jointly learns categories and their features. The set of features is shared across categories, and strength of association is inferred in a Bayesian framework. We approximate the learning environment with natural language text which allows us to evaluate performance on a large scale. Compared to highly engineered pattern-based approaches, our model is cognitively motivated, knowledge-lean, and learns categories and features which are perceived by humans as more meaningful.

Shared common ground influences information density in microblog texts

Gabriel Doyle and Michael Frank 14:25–14:50

If speakers use language rationally, they should structure their messages to achieve approximately uniform information density (UID), in order to optimize transmission via a noisy channel. Previous work identified a consistent increase in linguistic information across sentences in text as a signature of the UID hypothesis. This increase was derived from a predicted increase in context, but the context itself was not quantified. We use microblog texts from Twitter, tied to a single shared event (the baseball World Series), to quantify both linguistic and non-linguistic context. By tracking changes in contextual information, we predict and identify gradual and rapid changes in information content in response to in-game events. These findings lend further support to the UID hypothesis and highlights the importance of non-linguistic common ground for language production and processing.

Hierarchic syntax improves reading time prediction

Marten van Schijndel and William Schuler 14:50–15:15

Previous work has debated whether humans make use of hierarchic syntax when processing language (Frank and Bod, 2011; Fossum and Levy, 2012). This paper uses an eye-tracking corpus to demonstrate that hierarchic syntax significantly improves reading time prediction over a strong n-gram baseline. This study shows that an interpolated 5-gram baseline can be made stronger by combining n-gram statistics over entire eye-tracking regions rather than simply using the last n-gram in each region, but basic hierarchic syntactic measures are still able to achieve significant improvements over this improved baseline.

6 Workshops

]

Thursday–Friday

Thursday

Friday

Workshop 1: 1st Conference on Machine Translation

Organizers: *Barry Haddow, Philipp Koehn, Ondrej 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, Michaeel 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-kiu 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 Alkhouli, 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 Alkhouli, Andreas Guta, 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 Sudarikov, 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 Stymne, 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 Sudarikov, 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
  Marta 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, Wenyan 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, Chenhui 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
  Michal 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, Walid Aransa, Yaxing Wang, Marc Masana, Mercedes García-Martínez, Fethi 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 Etchegoyhen
- 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
  Aswarth Abhilash Dara and Yiu-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 Jakubícek, and Vojtech Kovár
- 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 Khristich, 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 Stymne
- 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 Barrault, 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 Alkhouli, Gabriel Bretschner, Jan-Thorsten Peter, Mohammed Hethnawi, Andreas Guta, 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
Fabienne 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 Cohan, Edward Grefenstette, Karl Moritz Hermann, Laura Rim ell, Jason Weston, and Scott Wen-tau Yih*

Venue: Governor's Square 12

Thursday, August 11, 2016

3:10–3:30 **Poster Session**

- Pair Distance Distribution: A Model of Semantic Representation
  Yonatan Ramni, Oded Maimon, and Evgeni Khmelnitsky
- Distilling Word Embeddings: An Encoding Approach
  Lili Mou, Ran Jia, Yan Xu, Ge Li, Lu Zhang, and Zhi Jin
- Learning Phone Embeddings for Word Segmentation of Child-Directed Speech
  Jianqiang Ma, Çağrı Çöltekin, and Erhard Hinrichs
- An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation
  Jey Han Lau and Timothy Baldwin
- Parameterized context windows in Random Indexing
  Tobias Norlund, David Nilsson, and Magnus Sahlgren
- Functional Distributional Semantics
  Guy Emerson and Ann Copestake
- Domain Adaptation for Neural Networks by Parameter Augmentation
  Yusuke Watanabe, Kazuma Hashimoto, and Yoshimasa Tsuruoka
- Making Sense of Word Embeddings
  Maria Pelevina, Nikolay Arefiev, Chris Biemann, and Alexander Panchenko
- Quantifying the Vanishing Gradient and Long Distance Dependency Problem in Recursive Neural Networks and Recursive LSTMs
  Phong Le and Willem Zuidema
- Assisting Discussion Forum Users using Deep Recurrent Neural Networks
  Jacob Hagstedt P Suorra and Olof Mogren
- Learning Text Similarity with Siamese Recurrent Networks
  Paul Neculoiu, Maarten Versteegh, and Mihai Rotaru
- Measuring Semantic Similarity of Words Using Concept Networks
  Gábor Recski, Eszter Iklódi, Katalin Pajkossy, and Andras Kornai
- Neural Associative Memory for Dual-Sequence Modeling
  Dirk Weissenborn
- Multilingual Modal Sense Classification using a Convolutional Neural Network
  Ana Marasović and Anette Frank
- Learning Semantic Relatedness in Community Question Answering Using Neural Models
  Henry Nassif, Mitra Mohtarami, and James Glass
- A Vector Model for Type-Theoretical Semantics
  Konstantin Sokolov

- Learning Word Importance with the Neural Bag-of-Words Model
  Imran Sheikh, Irina Illina, Dominique Fohr, and Georges Linarès
- Decomposing Bilexical Dependencies into Semantic and Syntactic Vectors
  Jeff Mitchell
- Learning Word Representations from Multiple Information Sources
  Yunchuan Chen, Lili Mou, Yan Xu, Ge Li, and Zhi Jin
- Explaining Predictions of Non-Linear Classifiers in NLP
  Leila Arras, Franziska Horn, Grégoire Montavon, Klaus-Robert Müller, and Wojciech Samek
- Mapping Unseen Words to Task-Trained Embedding Spaces
  Pranava Swaroop Madhyastha, Mohit Bansal, Kevin Gimpel, and Karen Livescu
- A Joint Model for Word Embedding and Word Morphology
  Kris Cao and Marek Rei
- Towards Generalizable Sentence Embeddings
  Eleni Triantafillou, Jamie Ryan Kiros, Raquel Urtasun, and Richard Zemel
- On the Compositionality and Semantic Interpretation of English Noun Compounds
  Corina Dima
- Using Embedding Masks for Word Categorization
  Stefan Ruseti, Traian Rebedea, and Stefan Trausan-Matu
- LSTM-Based Mixture-of-Experts for Knowledge-Aware Dialogues
  Phong Le, Marc Dymetman, and Jean-Michel Renders
- A Two-stage Approach for Extending Event Detection to New Types via Neural Networks
  Thien Huu Nguyen, Lisheng Fu, Kyunghyun Cho, and Ralph Grishman
- Joint Learning of Sentence Embeddings for Relevance and Entailment
  Petr Baudiš, Silvestr Stanko, and Jan Šedivý
- Towards cross-lingual distributed representations without parallel text trained with adversarial autoencoders
  Antonio Valerio Miceli Barone
- Sparsifying Word Representations for Deep Unordered Sentence Modeling
  Prasanna Sattigeri and Jayaraman J. Thiagarajan
- 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
- Why "Blow Out"? A Structural Analysis of the Movie Dialog Dataset
  Richard Searle and Megan Bingham-Walker

12:10–1:30 **Lunch Break**

Workshop 3: 10th Linguistic Annotation Workshop

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

Venue: Director's Row E

- Building a Cross-document Event-Event Relation Corpus
  Yu Hong, Tongtao Zhang, Tim O'Gorman, Sharone Horowit-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 Laarmann-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
- Different Flavors of GUM: Evaluating Genre and Sentence Type Effects on Multilayer Corpus Annotation Quality
  Amir Zeldes and Dan Simonson
- Addressing Annotation Complexity: The Case of Annotating Ideological Perspective in Egyptian Social Media
  Heba Elfardy 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 Srikumar, 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
- Dependency Annotation Choices: Assessing Theoretical and Practical Issues of Universal Dependencies
  Kim Gerdes and Sylvain Kahane
- Conversion from Paninian Karakas to Universal Dependencies for Hindi Dependency Treebank
  Juhi Tandon, Himani Chaudhry, Riyaz Ahmad Bhat, and Dipti Sharma

- 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
  Manjuan Duan, Ethan Hill, and Michael White
- Applying Universal Dependency to the Arapaho Language
  Irina Wagner, Andrew Cowell, and Jena D. Hwang
- Annotating the discourse and dialogue structure of SMS message conversations
  Nianwen Xue, Qishen Su, and Sooyoung Jeong
- Creating a Novel Geolocation Corpus from Historical Texts
  Grant DeLozier, Ben Wing, Jason Baldridge, and Scott Nesbit

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 Nokel 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
  Meghdad Farahmand and James Henderson
- Phrase Representations for Multiword Expressions
  Joël Legrand and Ronan Collobert
- Representing Support Verbs in FrameNet
  Miriam R L Petruck 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 Grigonyté, 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 Fermin Moscoso del Prado Martin

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 Reinflection
Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden

11:30–12:30 **Shared task poster session**

- Morphological reinflection with convolutional neural networks
  Robert Östling
- EHU at the SIGMORPHON 2016 Shared Task. A Simple Proposal: Grapheme-to-Phoneme for Inflection
  Iñaki Alegria and Izaskun Etxeberria
- Morphological Reinflection via Discriminative String Transduction
  Garrett Nicolai, Bradley Hauer, Adam St Arnaud, and Grzegorz Kondrak
- Morphological reinflection 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 Reinflection
  Roee 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 Reinflection
  Katharina Kann and Hinrich Schütze
- The Columbia University - New York University Abu Dhabi SIGMORPHON 2016 Morphological Reinflection 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
  Garrett 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 Zamaraeva

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
Estíbaliz 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 Hiippala
- 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 Etxeberria, Iñaki Alegria, Larraitz Uria, and Mans Hulden
- Towards Building a Political Protest Database to Explain Changes in the Welfare State
  Çağıl 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 Hockenmaier

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 Uslu, 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 Sheikhshab, 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, Sucheendra 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 Limsopatham 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, Dane 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 Kocbek 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 Mahanandeeshwar Gattu

Workshop 10: SIGFSM Workshop on Statistical NLP and Weighted Automata

Organizers: *Bryan Jurish, Andreas Maletti, Uwee 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 Forsberg 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, and Antoine Goldberg Yoav a bibnamedelimb nd 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 Chiu, Anna Korhonen, and Sampo Pyysalo
- A critique of word similarity as a method for evaluating distributional semantic models
  Miroslav Batchkarov, Thomas Kober, Jeremy Reffin, 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
  Allyson Ettinger and Tal Linzen
- Evaluating embeddings on dictionary-based similarity
  Judit Ács and Andras Kornai
- Evaluating multi-sense embeddings for semantic resolution monolingually and in word translation
  Gábor Borbély, Márton Makrai, Dávid Márk Nemeskey, and Andras 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
  Allyson 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

- Automatic Classification by Topic Domain for Meta Data Generation, Web Corpus Evaluation, and Corpus Comparison
  Roland Schäfer and Felix Bildhauer
- Efficient construction of metadata-enhanced web corpora
  Adrien Barbaresi
- Topically-focused Blog Corpora for Multiple Languages
  Andrew Salway, Dag Elgesem, Knut Hofland, Øystein Reigem, and Lubos Steskal
- The Challenges and Joys of Analysing Ongoing Language Change in Web-based Corpora: a Case Study
  Anne Krause
- Using the Web and Social Media as Corpora for Monitoring the Spread of Neologisms. The case of 'rapefugee', 'rapeugee', and 'rapugee'.
  Quirin Würschinger, Mohammad Fazleh Elahi, Desislava Zhekova, and Hans-Jörg Schmid
- Babler - Data Collection from the Web to Support Speech Recognition and Keyword Search
  Gideon Mendels, Erica Cooper, and Julia Hirschberg
- A Global Analysis of Emoji Usage
  Nikola Ljubešić and Darja Fišer
- Genre classification for a corpus of academic webpages
  Erika Dalan and Serge Sharoff
- On Bias-free Crawling and Representative Web Corpora
  Roland Schäfer
- EmpiriST 2015: A Shared Task on the Automatic Linguistic Annotation of Computer-Mediated Communication and Web Corpora
  Michael Beißwenger, Sabine Bartsch, Stefan Evert, and Kay-Michael Würzner
- SoMaJo: State-of-the-art tokenization for German web and social media texts
  Thomas Proisl and Peter Uhrig
- UdS-(retrain|distributional|surface): Improving POS Tagging for OOV Words in German CMC and Web Data
  Jakob Prange, Andrea Horbach, and Stefan Thater
- EmpiriST: AIPHES - Robust Tokenization and POS-Tagging for Different Genres
  Steffen Remus, Gerold Hintz, Chris Biemann, Christian M. Meyer, Darina Benikova, Judith Eckle-Kohler, Margot Mieskes, and Thomas Arnold
- bot.zen EmpiriST 2015 - A minimally-deep learning PoS-tagger (trained for German CMC and Web data)
  Egon Stemle
- LTL-UDE EmpiriST 2015: Tokenization and PoS Tagging of Social Media Text
  Tobias Horsmann and Torsten Zesch

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 Joakim 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 andPetasis, 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
Jaromir Savelka and Kevin D. Ashley

11:40–12:00 Scrutable Feature Sets for Stance Classification
Angrosh Mandya, 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**

7 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`

Index

8 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.

Art at Every Angle

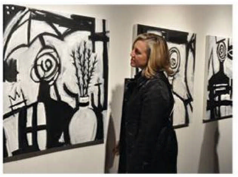
Exterior of Denver Art Museum by Jeff Wells, courtesy Denver Art Museum. Other photos by Steve Crocellius.

"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*," throught January 9, 2011.

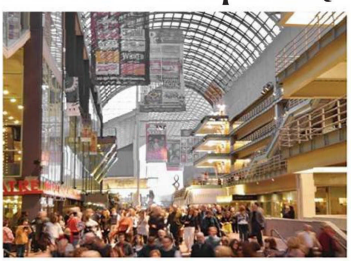
Inside, the 356,000-square-foot DAM is filled with classics by Monet, Picasso and Matisse, alongside more modern works by Warhol and O'Keefe. All in all, the museum contains a collection numbering more than 68,000 works from around the world, including intriguing pieces from Africa and pre-Columbian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

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** within 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.

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.

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.

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.

Denver 365
Denver's Year-Round Events Guide

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	71
FEBRUARY	46.6/20.2 F	8.1/-6.6 C	67/44	.57	70
MARCH	52.2/25.8 F	11.2/-3.4 C	68/40	1.28	69
APRIL	61.8/34.5 F	16.6/-1.4 C	67/35	1.71	67
MAY	70.8/43.6 F	21.6/6.4 C	70/38	2.40	65
JUNE	81.4/52.4 F	21.6/6.4 C	69/35	1.79	71
JULY	88.2/58.6 F	31.2/14.8 C	68/34	1.91	71
AUGUST	85.5/56.9 F	29.7/13.8 C	69/35	1.51	72
SEPTEMBER	76.9/47.6 F	24.9/8.7 C	69/34	1.24	74
OCTOBER	66.3/36.4 F	19.1/2.4 C	65/35	.98	72
NOVEMBER	52.5/25.4 F	11.4/-3.7 C	68/48	.87	65
DECEMBER	44.5/17.4 F	6.9/-8.1 C	65/51	.64	67
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.

Approximate Flight Times to/from Denver International Airport

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

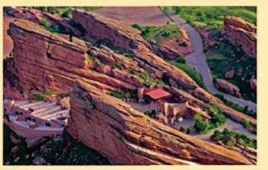
1 Stand a "mile high" on the steps of the **State Capitol** – 5,280 feet above sea level.

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 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.

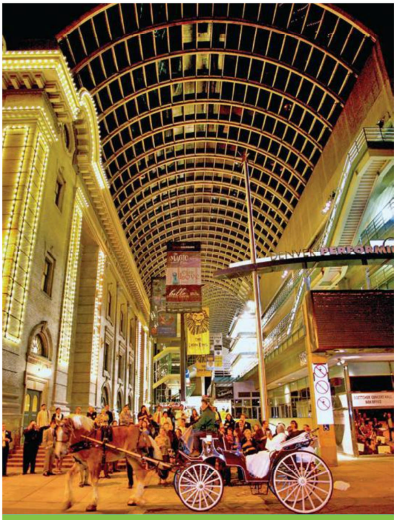
Free hybrid buses leave either end of the mall as often as every 90 seconds, stopping on every corner. The pedestrian portion of the 16th Street Mall continues over two footbridges to Commons Park on the South Platte River and to the Highlands neighborhood.

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 ¾-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.

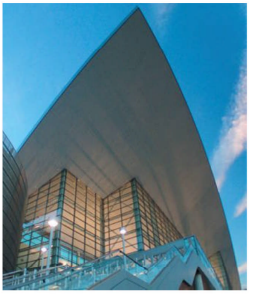
3 LARIMER SQUARE. Downtown Denver's trendiest block is also one of its most historic. Beautiful Victorian brick-and-stone buildings now house chic eateries, one-of-a-kind galleries and clothing boutiques, nightspots, a comedy club, and some of Denver's top restaurants. This is the perfect place to watch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

Top: 16th Street Mall (by Steve Cecchi)us. 1. D&F Tower by Rich Grant. 2. Writer's Square by Steve Cicoellius. 4. Denver Performing Arts Complex by Steve Cicoellius. 6. Colorado Convention Center by Steve Cicoellius. 7. Blue Bear by Rich Grant.) 11. Denver Pavilions by Steve Cicoellius.

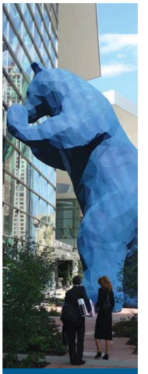


4 DENVER PERFORMING ARTS COMPLEX. The second-largest performing arts complex in the nation offers 10 venues seating more than 10,000 people for Tony® Award-winning theater, opera, symphony and ballet. The new $92-million Ellie Caulkins Opera House opened in 2005 and Boettcher Concert Hall is undergoing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

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.



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.



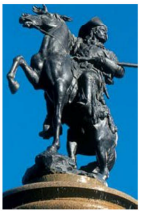
7 BLUE BEAR. One of the most popular public art statues in Denver is Lawrence Argent's 40-foot high Blue Bear, "*I See What You Mean*," which appears to be peeking into the convention center. The 10,000-pound bear is made up of 4,000 interlocking triangles.

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.

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	71
FEBRUARY	46.6/20.2 F	8.1/-6.6 C	67/44	.57	70
MARCH	52.2/25.8 F	11.2/-3.4 C	68/40	1.28	69
APRIL	61.8/34.5 F	16.6/-1.4 C	67/35	1.71	67
MAY	70.8/43.6 F	21.6/6.4 C	70/38	2.40	65
JUNE	81.4/52.4 F	21.6/6.4 C	69/35	1.79	71
JULY	88.2/58.6 F	31.2/14.8 C	68/34	1.91	71
AUGUST	85.5/56.9 F	29.7/13.8 C	69/35	1.51	72
SEPTEMBER	76.9/47.6 F	24.9/8.7 C	69/34	1.24	74
OCTOBER	66.3/36.4 F	19.1/2.4 C	65/35	.98	72
NOVEMBER	52.5/25.4 F	11.4/-3.7 C	68/48	.87	65
DECEMBER	44.5/17.4 F	6.9/-8.1 C	65/51	.64	67
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.

Approximate Flight Times to/from Denver International Airport

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

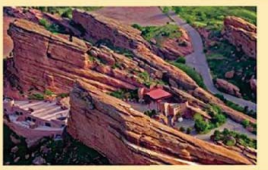
1 Stand a "mile high" on the steps of the **State Capitol** – 5,280 feet above sea level.

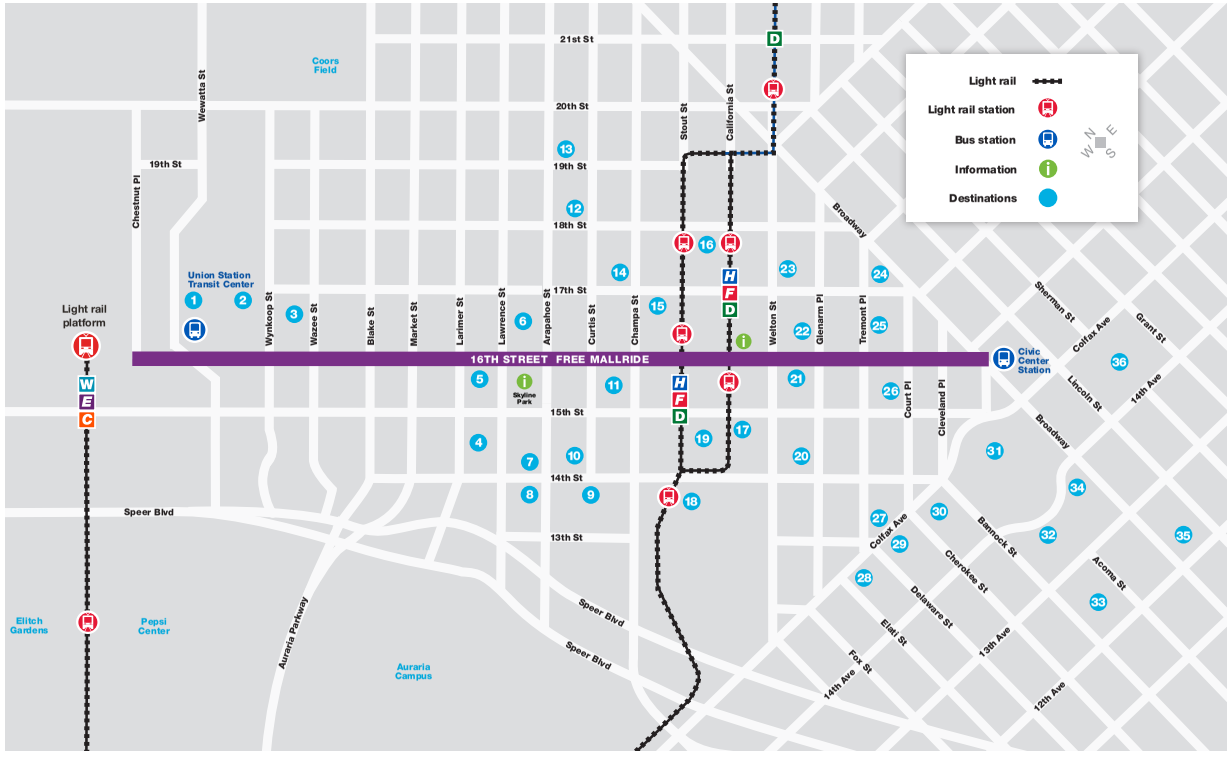
2 Ride the FREE hybrid shuttle bus on the mile-long **16th Street Mall**.

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.





1. Union Station Transit Center
2. The Crawford Hotel
3. Oxford Hotel
4. Larimer Square
5. Writer Square
6. Westin Hotel
7. Four Seasons Hotel
8. Hotel Teatro
9. Denver Performing Arts Complex
10. Curtis Hotel
11. Courtyard Marriott
12. Ritz Carlton
13. Greyhound Bus Terminal
14. Hotel Monaco
15. Magnolia Hotel
16. Denver Marriott City Center
17. Hyatt Regency Hotel
18. Colorado Convention Center
19. Hilton Garden Inn
20. Crowne Plaza Hotel
21. Denver Pavilions
22. Paramount Theater
23. Grand Hyatt Denver
24. Brown Palace Hotel
25. Republic Plaza
26. Sheraton Downtown Denver Hotel
27. Firefighters Museum
28. Denver Justice Center Campus
29. U.S. Mint
30. City Hall/City & County Building
31. Civic Center Park
32. Denver Art Museum
33. Byers-Evans House Museum
34. Denver Central Library
35. History Colorado
36. State Capital

NAACL gratefully acknowledges the following sponsors for their support:

Platinum

Bloomberg

Gold

Google Goldman Sachs A9 IBM Watson

Silver

NUANCE SDL Research

Bronze

YAHOO! LABS

Supporter

digital roots UNIVERSITY of WASHINGTON Computational Linguistics

Student Research Workshop

NSF Google BAOBAB fuse | machines

Wireless

MACHINE ZONE