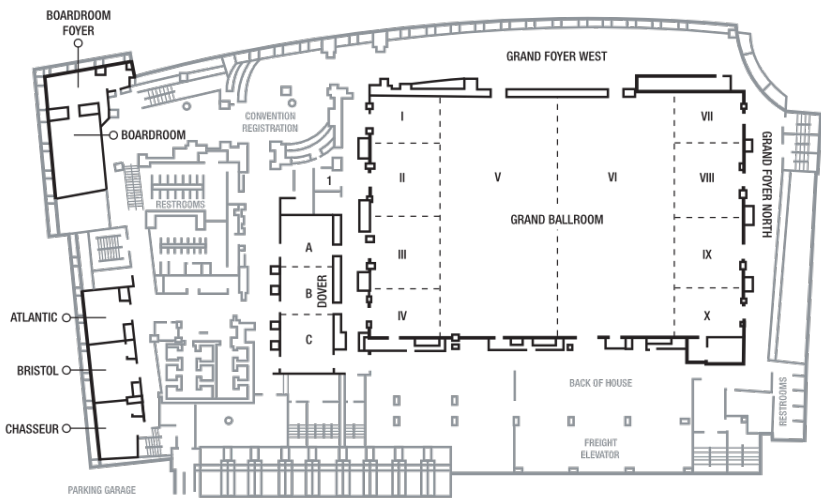The 52nd Annual Meeting of the Association for Computational Linguistics
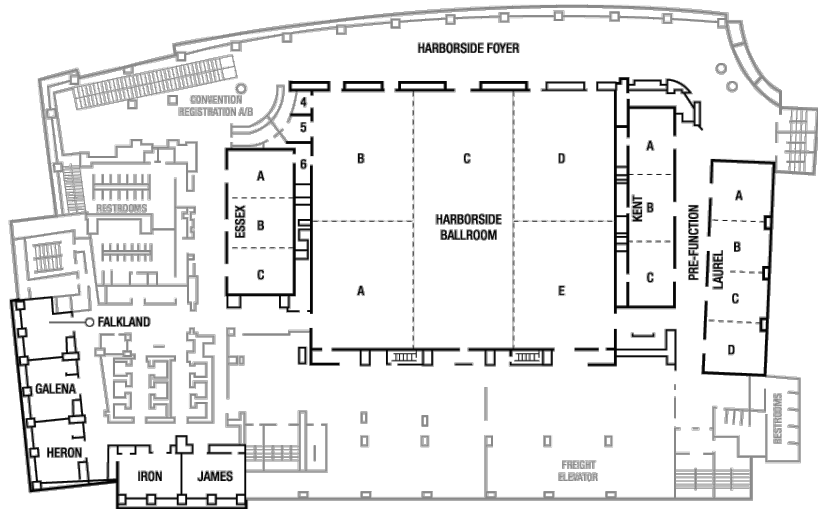
Conference Handbook

ACL 2014
June 22–27
Baltimore

THIRD FLOOR LEVEL

BOARDROOM FOYER

BOARDROOM

CONVENTION REGISTRATION

GRAND FOYER WEST

GRAND FOYER NORTH

RESTROOMS

1

I
II
III
IV

V

VI

VII
VIII
IX
X

GRAND BALLROOM

A
B
C

DOVER

ATLANTIC

BRISTOL

CHASSEUR

BACK OF HOUSE

RESTROOMS

FREIGHT ELEVATOR

PARKING GARAGE

FOURTH FLOOR LEVEL

HARBORSIDE FOYER

CONVENTION REGISTRATION A/B

RESTROOMS

4
5
6

A
B
C

ESSEX

B

C

D

A

E

A
B
C

KENT

A
B
C
D

PRE-FUNCTION

LAUREL

HARBORSIDE BALLROOM

FALKLAND

GALENA

HERON

IRON

JAMES

RESTROOMS

FREIGHT ELEVATOR

Contents

i

Conference Information

Message from the General Chair

Welcome to ACL 2014!

I remember with great fondness the first ACL Conference I attended 20 years ago in Las Cruces, New Mexico. Some things have changed: papers presented there that I considered interesting or inconsequential have switched positions in my personal ranking as I learned more and more about our field; single sessions have long been replaced by parallel sessions to accommodate an ever increasing number of research contributions; the number of associated workshops and posters has mushroomed beyond anyone's dream. Almost without noticing, we transitioned from small conferences of a few hundred to conferences that bring together 1000 plus participants from all over the world. Our field has matured significantly attracting the attention of not only a handful of academics, but successful industries and Research Labs as well. Some things have stayed the same though: ACL continues to be the pre-eminent conference in our field and the best place to meet and make like-minded friends, discuss tantalizing tricks that you can learn about only in face-to-face communication settings, and celebrate the results we get.

On behalf of the organizing committee, I welcome you to the conference; make the most of it!

Daniel Marcu
General Chair

Message from the Program Committee Co-Chairs

Welcome to the 2014 Conference of the Association for Computational Linguistics! This year ACL received 572 long paper submissions and 551 short paper submissions. Of the long papers, 146 were accepted for presentation at ACL — 95 as oral, and 51 as poster presentations. 139 short papers were accepted — 51 as oral, and 88 as poster presentations.

The submissions were reviewed under different categories and using different review forms for empirical/data-driven, theoretical, applications/tools, resources/evaluation, and survey papers. For the short papers we additionally used a negative results category and were glad to see that the community is becoming more open to enabling the publication of useful negative results.

Based on feedback from prior years, this year we organized the posters in two large poster sessions to accommodate the growing number of high-quality submissions accepted in poster presentation format. We hope attendees and authors will benefit from this additional time to present and discuss ideas. Another innovation we are experimenting with this year is to optimize the conference schedule based on feedback from attendees on the talks they would like to see. We collected attendee responses using a scheduling survey developed with the help of David Yarowsky and Svitlana Volkova (thanks to the 338 volunteers who completed the survey!), and we optimized the conference schedule to assign popular sessions to large conference rooms, and to reduce the chance that two talks that an attendee is interested in are scheduled at the same time. Additionally, as in NAACL 2013, all talks will be recorded and made available for future viewing.

ACL 2014 will have two distinguished invited speakers. Corinna Cortes (Head of Google Research, NY) and Zoran Popović (Director of Center for Game Science, University of Washington).

There are many individuals to thank for their contributions to ACL 2014. We would like to thank the thirty three area chairs for their hard work on recruiting reviewers, leading the discussion process, and carefully ranking the submissions. We would like to thank Mark Dredze for developing and sharing a reviewer assignment tool, that was used at ACL this year. It was applied to ACL reviewing with the help of Jiang Guo and the area chairs who provided feedback at several stages of the process. We would also like to thank the seven hundred and seventy nine reviewers and seventy two secondary reviewers on whose efforts we depend to select high-quality and timely scientific work. This year we specifically acknowledged around 14% of the reviewers who went the extra mile and provided extremely helpful to the area chairs and authors reviews (their names are marked with a * in the organization section of the proceedings). The ACL coordinating committee members, including Dragomir Radev, Jian Su, Graeme Hirst, Hal Daumé III, Chris Callison-Burch, and Haifeng Wang were very helpful on various issues relating to the organization. We would like to thank the prior conference chairs Jason Eisner, Hal Daumé, Lucy Vanderwende, Jian Su, Rada Mihalcea, Mar-

ius Pasca, Pascale Fung, and Massimo Poesio for their advice. We are very grateful for the guidance and support of the general chair Daniel Marcu, to the ACL Business Manager Priscilla Rasmussen who knew practically everything, to the local chair David Yarowsky, the publication chairs Yusuke Miyao and Alexander Koller, and to Matt Post who stepped in to handle the conference handbook. We would also like to thank Jiang Guo who helped with reviewer assignment and numerous other tasks. Rich Gerber from Softconf was extremely responsive to all of our requests, and we are grateful for that.

We hope you will enjoy ACL 2014 in Baltimore!

ACL 2014 Program Co-Chairs
Kristina Toutanova, Microsoft Research
Hua Wu, Baidu

Conference Committee

General Conference Chair
Daniel Marcu, ISI/USC

Program Committee Chairs
Kristina Toutanova, Microsoft Research
Hua Wu, Baidu

Local Arrangements Chair
David Yarowsky, Johns Hopkins University

Workshop Chairs
Jill Burstein, Educational Testing Service
Lluís Màrquez, Universitat Politècnica de Catalunya

Tutorial Chairs
Alex Fraser, University of Munich
Yang Liu, Tsinghua University

Publications Chairs
Alexander Koller, University of Potsdam
Yusuke Miyao, National Institute of Informatics Japan

Student Research Workshop Chairs
Ekaterina Kochmar, University of Cambridge
Annie Louis, University of Pennsylvania
Svitlana Volkova, Johns Hopkins University

Faculty Advisors for the Student Research Workshop
Bill Byrne, University of Cambridge
Jordan Boyd-Graber, University of Maryland

Demonstrations Chairs
Kalina Bontcheva, University of Sheffield
Jingbo Zhu, Northeastern University (China)

Publicity Chair
Jason Riesa, Google

Business Manager
Priscilla Rasmussen

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

Area Chairs
Klinton Bicknell, Northwestern University
Asli Celikyilmaz, Microsoft Research
Micha Elsner, Ohio State University
Michael Strube, HITS gGmbH
Jacob Eisenstein, Georgia Institute of Technology
Jun Zhao, Chinese Academy of Sciences
Barbara Di Eugenio, University of Illinois at Chicago
Ting Liu, Harbin Institute of Technology

Marius Pasca, Google Research
Mihai Surdeanu, University of Arizona
Nicoletta Calzolari, Institute of Computational Linguistics of the National Research Council
Nianwen Xue, Brandeis University
Eneko Agirre, University of the Basque Country
Kevin Duh, Nara Institute of Science and Technology
Jason Eisner, Johns Hopkins University
Colin Cherry, National Research Council, Canada
Niyu Ge, IBM Research
Liang Huang, City University of New York
Yanjun Ma, Baidu
Michel Galley, Microsoft Research
Brian Roark, Google Research
Mamoru Komachi, Tokyo Metropolitan University
Miles Osborne, Johns Hopkins University
Alan Ritter, Carnegie Mellon University
Iryna Gurevych, Technische Universität Darmstadt
Percy Liang, Stanford University
Shiqi Zhao, Baidu
Yang Liu, University of Texas at Dallas
Scott Wen-tau Yih, Microsoft Research
Ciprian Chelba, Google Research
Xavier Carreras, Universitat Politecnica de Catalunya
Slav Petrov, Google Research
Grzegorz Kondrak, University of Alberta

Conference Handbook Editor
Matt Post, Johns Hopkins University

Local Guide Editors
Ann Irvine, Johns Hopkins University
Juri Ganitkevitch, Johns Hopkins University

Student Volunteer Chair
Yuan Cao, Johns Hopkins University

Entertainment Chairs
Naomi Saphra, Johns Hopkins University
Matt Gormley, Johns Hopkins University

Graphic Design
Jeff Hayes, `Jeff-Hayes.com`
Meg Mitchell, Microsoft Research
Keisuke Sakaguchi, Johns Hopkins University

Local JHU/CLSP Administration
Laura Graham
Ruth Scally

2

Tutorials: Sunday, June 22

Overview

Syntactic Processing Using Global Discriminative Learning and Beam-Search
Decoding   (Salon IV)
Yue Zhang, Meishan Zhang, and Ting Liu

6:00 – 9:00   **Welcome Reception**   (Grand Ballroom)

Tutorial 1

Gaussian Processes for Natural Language Processing

Trevor Cohn (University of Melbourne), Daniel Preoţiuc-Pietro (University of Sheffield), Neil Lawrence (University of Sheffield)

Sunday, June 22, 2014, 9:00 – 12:30pm

Salons I–II

This tutorial aims to cover the basic motivation, ideas and theory of Gaussian Processes and several applications to natural language processing tasks. Gaussian Processes (GPs) are a powerful modelling framework incorporating kernels and Bayesian inference, and are recognised as state-of-the-art for many machine learning tasks. This tutorial will focus primarily on regression and classification, both fundamental techniques of wide-spread use in the NLP community. We argue that the GP framework offers many benefits over commonly used machine learning frameworks, such as linear models (logistic regression, least squares regression) and support vector machines (SVMs). GPs have the advantage of being a fully Bayesian model, giving a posterior over the desired variables. Their probabilistic formulation allows for much wider applicability in larger graphical models, unlike SVMs. Moreover, several properties of Gaussian distributions means that GP (regression) supports analytic formulations for the posterior and predictive inference, avoiding the many approximation errors that plague approximate inference techniques in common use for Bayesian models (e.g. MCMCM, variational Bayes).

GPs provide an elegant, flexible and simple means of probabilistic inference. GPs have been actively researched since the early 2000s, and are now reaching maturity: the fundamental theory and practice is well understood, and now research is focused into their applications, and improve inference algorithms, e.g. for scaling inference to large and high-dimensional datasets. Several open-source packages (e.g. GPy and GPML) have been developed which allow for GPs to be easily used for many applications. This tutorial aims to present the main ideas and theory behind GPs and recent applications to NLP, emphasising their potential for widespread application across many NLP tasks.

Trevor Cohn is a Senior Lecturer and ARC Future Fellow at the University of Melbourne. His research deals with probabilistic machine learning models, particularly structured prediction and non-parametric Bayesian models. He has recently published several seminal papers on Gaussian Process models for NLP with applications ranging from translation evaluation to temporal dynamics in social media.

Daniel Preoţiuc-Pietro is a Research Associate at the University of Sheffield. His research involves applying machine learning models to model large volumes of data, usually from social media. Applications include forecasting future behaviours of text, users or real world quantities (e.g. political voting intention), user geo-location and impact.

Neil Lawrence is a Professor at the University of Sheffield. He is one of the foremost experts on Gaussian Processes and non-parametric Bayesian inference, with a long history of publications and innovations in the field, including their application to multi-output scenarios, unsupervised learning, deep networks and scaling to big data. He has been program chair for top machine learning conferences (NIPS, AISTATS), and has run several past tutorials on Gaussian Processes.

Tutorial 2

Scalable Large-Margin Structured Learning: Theory and Algorithms

Liang Huang (CUNY), Kai Zhao (CUNY), and Lemao Liu (CUNY)

Sunday, June 22, 2014, 9:00 – 12:30pm

Salons VII–VIII

Much of NLP tries to map structured input (sentences) to some form of structured output (tag sequences, parse trees, semantic graphs, or translated/paraphrased/compressed sentences). Thus structured prediction and its learning algorithm are of central importance to us NLP researchers. However, when applying machine learning to structured domains, we often face scalability issues for two reasons:

1. Even the fastest exact search algorithms for most NLP problems (such as parsing and translation) is too slow for repeated use on the training data, but approximate search (such as beam search) unfortunately breaks down the nice theoretical properties (such as convergence) of existing machine learning algorithms.

2. Even with inexact search, the scale of the training data in NLP still makes pure online learning (such as perceptron and MIRA) too slow on a single CPU.

This tutorial reviews recent advances that address these two challenges. In particular, we will cover principled machine learning methods that are designed to work under vastly inexact search, and parallelization algorithms that speed up learning on multiple CPUs. We will also extend structured learning to the latent variable setting, where in many NLP applications such as translation and semantic parsing the gold-standard derivation is hidden.

Liang Huang is an Assistant Professor at the City University of New York (CUNY). He graduated in 2008 from Penn and has worked as a Research Scientist at Google and a Research Assistant Professor at USC/ISI. His work is mainly on the theoretical aspects (algorithms and formalisms) of computational linguistics, as well as theory and algorithms of structured learning. He has received a Best Paper Award at ACL 2008, several best paper nominations (ACL 2007, EMNLP 2008, and ACL 2010), two Google Faculty Research Awards (2010 and 2013), and a University Graduate Teaching Prize at Penn (2005). He has given two tutorials at COLING 2008 and NAACL 2009, being the most popular tutorial at both venues.

Kai Zhao is a Ph.D. candidate at the City University of New York (CUNY), working with Liang Huang. He received his B.S. from the University of Science and Technology in China (USTC). He has published on structured prediction, online learning, machine translation, and parsing algorithms. He was a summer intern with IBM TJ Watson Research Center in 2013.

Lemao Liu is a postdoctoral research associate at the City University of New York (CUNY), working with Liang Huang. He received his Ph.D. from the Harbin Institute of Technology in 2013. Much of his Ph.D. work was done while visiting NICT, Japan, under the guidance of Taro Watanabe. His research area is machine translation and machine learning.

Tutorial 3

Semantics for Large-Scale Multimedia: New Challenges for NLP

Florian Metze (CMU) and Koichi Shinoda (TokyoTech)

Sunday, June 22, 2014, 9:00 – 12:30pm

Salon III

Thousands of videos are constantly being uploaded to the web, creating a vast resource, and an ever-growing demand for methods to make them easier to retrieve, search, and index. As it becomes feasible to extract both low-level as well as high-level (symbolic) audio, speech, and video features from this data, these need to be processed further, in order to learn and extract meaningful relations between these. The language processing community has made huge process in analyzing the vast amounts of very noisy text data that is available on the Internet. While it is very difficult to create semantic units of low-level image descriptors or non-speech sounds by themselves, it is comparatively easy to ground semantics in the word output of a speech recognizer, or text data that is loosely associated with a video. This creates an opportunity for NLP researchers to use their unique skills, and make significant contributions to solve tasks on data that is even noisier than web text, but (we argue) even more interesting and challenging.

This tutorial aims to present to the NLP community the state of the art in audio and video processing, by discussing the most relevant tasks at NIST's TREC Video Retrieval Evaluation (TRECVID) workshop series. We liken "Semantic Indexing" (SIN) task, in which a system must identify occurrences of concepts such as "desk", or "dancing" in a video to the word spotting approach. We then proceed to explain more recent, and challenging tasks, "Multimedia Event Detection" (MED) and "Multimedia Event Recounting" (MER), which can be compared to transcription and summarization tasks. Finally, we will present an easy way to get started in multi-media analysis using Virtual Machines from the "Speech Recognition Virtual Kitchen", which will enable tutorial participants to perform hands-on experiments during the tutorial, and at home.

Florian Metze received his PhD from Universitat Karlsruhe (TH) in 2005. He worked as a Senior Research Scientist at Deutsche Telekom Laboratories (T-Labs) and joined Carnegie Mellon University's faculty in 2009. His interests includes speech and audio processing, and user interfaces.

Koichi Shinoda received his D. Eng. from Tokyo Institute of Technology in 2001. In 1989, he joined NEC Corporation. From 1997 to 1998, he was a visiting scholar with Bell Labs, Lucent Technologies. He is currently a Professor at the Tokyo Institute of Technology. His research interests include speech recognition, video information retrieval, and human interfaces.

Tutorial 4

Wikification and Beyond: The Challenges of Entity and Concept Grounding

Dan Roth (UIUC), Heng Ji (RPI), Ming-Wei Chang (MSR), Taylor Cassidy (ARL, IBM)

Sunday, June 22, 2014, 9:00 – 12:30pm

Dover A/B/C

Contextual disambiguation and grounding of concepts and entities in natural language text are essential to moving forward in many natural language understanding related tasks and are fundamental to many applications. The Wikification task aims at automatically identifying concept mentions appearing in a text document and linking them to (or "grounding them in") concept referents in a knowledge base (KB) (e.g., Wikipedia). For example, consider the sentence, "The Times report on Blumenthal (D) has the potential to fundamentally reshape the contest in the Nutmeg State.". A Wikifier should identify the key entities and concepts (Times, Blumental, D and the Nutmeg State), and disambiguate them by mapping them to an encyclopedic resource revealing, for example, that "D" here represents the Democratic Party, and that "the Nutmeg State" refers Connecticut. Wikification may benefit both human end-users and Natural Language Processing (NLP) systems. When a document is Wikified a reader can more easily comprehend it, as information about related topics and relevant enriched knowledge from a KB is readily accessible. From a system-to-system perspective, a Wikified document conveys the meanings of its key concepts and entities by grounding them in an encyclopedic resource or a structurally rich ontology.

The primary goals of this tutorial are to review the framework of Wikification and motivate it as a broad paradigm for cross-source linking for knowledge enrichment. We will present and

---

Dan Roth is a Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign and the Beckman Institute of Advanced Science and Technology (UIUC) and a University of Illinois Scholar. He is a fellow of AAAI, ACL and the ACM. Roth has published broadly in machine learning, natural language processing, knowledge representation and reasoning and received several paper, teaching and research awards. He has developed several machine learning based natural language processing systems that are widely used in the computational linguistics community and in industry and has presented invited talks and tutorials in several major conferences. Over the last few years he has worked on Entity Linking and Wikification. He has given several tutorials at ACL/NAACL/ECL and other forums.

Heng Ji is the Edward G. Hamilton Development Chair Associate Professor in Computer Science Department of Rensselaer Polytechnic Institute. Her research interests focus on Natural Language Processing, especially on Cross-source Information Extraction and Knowledge Base Population. She coordinated the NIST TAC Knowledge Base Population task in 2010, 2011 and 2014 and has published several papers on entity linking and Wikification.

Ming-Wei Chang is a researcher at Microsoft Research. His research interests are in machine learning and natural language understanding. He currently focuses on using large-scale structured and unstructured data for semantic understanding. Specially, he is interested in developing algorithms for entity linking that are effective for short and noisy text.

Taylor Cassidy is a Postdoctoral Researcher at U.S. Army Research Laboratory & IBM Research. His research interests include Cross-lingual Entity Linking and Wikification for social media.

discuss multiple dimensions of the task definition, present the basic building blocks of a state-of-the-art Wikifier system, share some key lessons learned from the analysis of evaluation results, and discuss recently proposed ideas for advancing work in this area along with some of the key challenges. We will also suggest some research questions brought up by new applications, including interactive Wikification, social media, and censorship. The tutorial will be useful for both senior and junior researchers with interests in cross-source information extraction and linking, knowledge acquisition, and the use of acquired knowledge in natural language processing and information extraction. We will try to provide a concise roadmap of recent perspectives and results, as well as point to some of our Wikification resources that are available to the research communities.

Tutorial 5

New Directions in Vector Space Models of Meaning

Phil Blunsom (Oxford), Georgiana Dinu (Trento), Edward Grefenstette (Oxford), and Karl Moritz Hermann (Oxford)

Sunday, June 22, 2014, 2:00 – 5:30pm

Salons I–III

Symbolic approaches have dominated NLP as a means to model syntactic and semantic aspects of natural language. While powerful inferential tools exist for such models, they suffer from an inability to capture correlation between words and to provide a continuous model for word, phrase, and document similarity. Distributed representations are one mechanism to overcome these constraints.

This tutorial will supply NLP researchers with the mathematical and conceptual background to make use of vector-based models of meaning in their own research. We will begin by motivating the need for a transition from symbolic representations to distributed ones. We will briefly cover how collocational (distributional) vectors can be used and manipulated to model word meaning. We will discuss the progress from distributional to distributed representations, and how neural networks allow us to learn word vectors and condition them on metadata such as parallel texts, topic labels, or sentiment labels. Finally, we will present various forms of semantic vector composition, and discuss their relative strengths and weaknesses, and their application to problems such as language modelling, paraphrasing, machine translation and document classification.

This tutorial aims to bring researchers up to speed with recent developments in this fast-moving field. It aims to strike a balance between providing a general introduction to vector-based models of meaning, an analysis of diverging strands of research in the field, and also being a hands-on tutorial to equip NLP researchers with the necessary tools and background knowledge to start working on such models. Attendees should be comfortable with basic probability, linear algebra, and continuous mathematics. No substantial knowledge of machine learning is required.

Phil Blunsom is an Associate Professor at the University of Oxford's Department of Computer Science. His research centres on the probabilistic modelling of natural languages, with a particular interest in automating the discovery of structure and meaning in text.

Georgiana Dinu is a postdoctoral researcher at the University of Trento. Her research revolves around distributional semantics with a focus on compositionality within the distributional paradigm.

Edward Grefenstette is a postdoctoral researcher at Oxford's Department of Computer Science. He works on the relation between vector representations of language meaning and structured logical reasoning. His work in this area was recently recognised by a best paper award at *SEM 2013.

Karl Moritz Hermann is a final-year DPhil student at the Department of Computer Science in Oxford. His research studies distributed and compositional semantics, with a particular emphasis on mechanisms to reduce task-specific and monolingual syntactic bias in such representations.

Tutorial 6

Structured Belief Propagation for NLP

Matthew R. Gormley (JHU), Jason Eisner (JHU)

Sunday, June 22, 2014, 2:00 – 5:30pm

Dover A/B/C

Statistical natural language processing relies on probabilistic models of linguistic structure. More complex models can help capture our intuitions about language, by adding linguistically meaningful interactions and latent variables. However, inference and learning in the models we want often poses a serious computational challenge. Belief propagation (BP) and its variants provide an attractive approximate solution, especially using recent training methods. These approaches can handle joint models of interacting components, are computationally efficient, and have extended the state-of-the-art on a number of common NLP tasks, including dependency parsing, modeling of morphological paradigms, CCG parsing, phrase extraction, semantic role labeling, and information extraction (Smith and Eisner, 2008, Dreyer and Eisner, 2009, Auli and Lopez, 2011, Burkett and Klein, 2012, Naradowsky et al., 2012, Stoyanov and Eisner, 2012).

This tutorial delves into BP with an emphasis on recent advances that enable state-of-the-art performance in a variety of tasks. Our goal is to elucidate how these approaches can easily be applied to new problems. We also cover the theory underlying them. Our target audience is researchers in human language technologies; we do not assume familarity with BP. In the first three sections, we discuss applications of BP to NLP problems, the basics of modeling with factor graphs and message passing, and the theoretical underpinnings of "what BP is doing" and how it relates to other variational inference techniques. In the second three sections, we cover key extensions to the standard BP algorithm to enable modeling of linguistic structure, efficient inference, and approximation-aware training. We survey a variety of software tools and introduce a new software framework that incorporates many of the modern approaches covered in this tutorial.

Matt Gormley is a PhD student at Johns Hopkins University working with Mark Dredze and Jason Eisner. His current research focuses on joint modeling of multiple linguistic strata in learning settings where supervised resources are scarce. He has authored papers in a variety of areas including topic modeling, global optimization, semantic role labeling, and grammar induction.

Jason Eisner is an Associate Professor in Computer Science and Cognitive Science at Johns Hopkins University, where he has received two school-wide awards for excellence in teaching. His 80+ papers have presented many models and algorithms spanning numerous areas of NLP. His goal is to develop the probabilistic modeling, inference, and learning techniques needed for a unified model of all kinds of linguistic structure. In particular, he and his students introduced structured belief propagation, which integrates classical NLP models and their associated dynamic programming algorithms, as well as loss-calibrated training for use with belief propagation.

Tutorial 7

Semantics, Discourse and Statistical Machine Translation

Deyi Xiong (Sochoow University), Min Zhang (Sochoow University)

Sunday, June 22, 2014, 2:00 – 5:30pm

Salons VII–VIII

In the past decade, statistical machine translation (SMT) has been advanced from word-based SMT to phrase- and syntax-based SMT. Although this advancement produces significant improvements in BLEU scores, crucial meaning errors and lack of cross-sentence connections at discourse level still hurt the quality of SMT-generated translations. More recently, we have witnessed two active movements in SMT research: one towards combining semantics and SMT in attempt to generate not only grammatical but also meaning-preserved translations, and the other towards exploring discourse knowledge for document-level machine translation in order to capture inter-sentence dependencies.

The emergence of semantic SMT are due to the combination of two factors: the necessity of semantic modeling in SMT and the renewed interest of designing models tailored to relevant NLP/SMT applications in the semantics community. The former is represented by recent numerous studies on exploring word sense disambiguation, semantic role labeling, bilingual semantic representations as well as semantic evaluation for SMT. The latter is reflected in CoNLL shared tasks, SemEval and SenEval exercises in recent years.

The need of capturing cross-sentence dependencies for document-level SMT triggers the resurgent interest of modeling translation from the perspective of discourse. Discourse phenomena,

---

Dr. Deyi Xiong is a professor at Sochoow University. His research interests are in the area of natural language processing, particularly statistical machine translation and parsing. Previously he was a research scientist at the Institute for Infocomm Research of Singapore. He received the B.Sc degree from China University of Geosciences (Wuhan, China) in 2002, the Ph.D.degree from the Institute of Computing Technology (Beijing, China) in 2007, both in computer science. He has published papers in prestigious journals and conferences on statistical machine translation, including Computational Linguistics, IEEE TASLP, JAIR, NLE, ACL, EMNLP, AAAI and IJCAI. He was the program co-chair of IALP 2012 and CLIA workshop 2011.

Dr. Min Zhang, a distinguished professor and Director of the Research Center of Human Language Technology at Soochow University (China), received his Bachelor degree and Ph.D. degree in computer science from Harbin Institute of Technology in 1991 and 1997, respectively. From 1997 to 1999, he worked as a postdoctoral research fellow in Korean Advanced Institute of Science and Technology in South Korea. He began his academic and industrial career as a researcher at Lernout & Hauspie Asia Pacific (Singapore) in Sep. 1999. He joined Infotalk Technology (Singapore) as a researcher in 2001 and became a senior research manager in 2002. He joined the Institute for Infocomm Research (Singapore) as a research scientist in Dec. 2003. He joined the Soochow University as a distinguished professor in 2012.

His current research interests include machine translation, natural language processing, information extraction, social network computing and Internet intelligence. He has co-authored more than 150 papers in leading journals and conferences, and co-edited 10 books/proceedings published by Springer and IEEE. He was the recipient of several awards in China and oversea. He is the vice president of COLIPS (2011-2013), the elected vice chair of SIGHAN/ACL (2014-2015), a steering committee member of PACLIC (2011-now), an executive member of AFNLP (2013-2014) and a member of ACL (since 2006). He supervises Ph.D students at National University of Singapore, Harbin Institute of Technology and Soochow University.

such as coherent relations, discourse topics, lexical cohesion that are beyond the scope of conventional sentence-level n-grams, have been recently considered and explored in the context of SMT.

This tutorial aims at providing a timely and combined introduction of such recent work along these two trends as discourse is inherently connected with semantics. The tutorial has three parts. The first part critically reviews the phrase- and syntax-based SMT. The second part is devoted to the lines of research oriented to semantic SMT, including a brief introduction of semantics, lexical and shallow semantics tailored to SMT, semantic representations in SMT, semantically motivated evaluation as well as advanced topics on deep semantic learning for SMT. The third part is dedicated to recent work on SMT with discourse, including a brief review on discourse studies from linguistics and computational viewpoints, discourse research from monolingual to multilingual, discourse-based SMT and a few advanced topics.

The tutorial is targeted for researchers in the SMT, semantics and discourse communities. In particular, the expected audience comes from two groups: 1) Researchers and students in the SMT community who want to design cutting-edge models and algorithms for semantic SMT with various semantic knowledge and representations, and who would like to advance SMT from sentence-by-sentence translation to document-level translation with discourse information; 2) Researchers and students from the semantics and discourse community who are interested in developing models and methods and adapting them to SMT.

Tutorial 8

Incremental Structured Prediction Using a Global Learning and Beam-Search Framework

Yue Zhang (SUTD), Meishan Zhang (HIT-SCIR), Ting Liu (HIT-SCIR)

Sunday, June 22, 2014, 2:00 – 5:30pm

Salon IV

In the past decade, statistical machine translation (SMT) has been advanced from word-based SMT to phrase- and syntax-based SMT. Although this advancement produces significant improvements in BLEU scores, crucial meaning errors and lack of cross-sentence connections at discourse level still hurt the quality of SMT-generated translations. More recently, we have witnessed two active movements in SMT research: one towards combining semantics and SMT in attempt to generate not only grammatical but also meaning-preserved translations, and the other towards exploring discourse knowledge for document-level machine translation in order to capture inter-sentence dependencies.

The emergence of semantic SMT are due to the combination of two factors: the necessity of semantic modeling in SMT and the renewed interest of designing models tailored to relevant NLP/SMT applications in the semantics community. The former is represented by recent numerous studies on exploring word sense disambiguation, semantic role labeling, bilingual semantic representations as well as semantic evaluation for SMT. The latter is reflected in CoNLL shared tasks, SemEval and SenEval exercises in recent years.

The need of capturing cross-sentence dependencies for document-level SMT triggers the resurgent interest of modeling translation from the perspective of discourse. Discourse phenomena, such as coherent relations, discourse topics, lexical cohesion that are beyond the scope of conventional sentence-level n-grams, have been recently considered and explored in the context of SMT.

This tutorial aims at providing a timely and combined introduction of such recent work along these two trends as discourse is inherently connected with semantics. The tutorial has three parts. The first part critically reviews the phrase- and syntax-based SMT. The second part is devoted

---

Yue Zhang is an Assistant Professor at Singapore University of Technology and Design (SUTD). Before joining SUTD in 2012, he worked as a postdoctoral research associate at University of Cambridge. He received his PhD and MSc degrees from University of Oxford, and undergraduate degree from Tsinghua University, China. Dr Zhang's research interest includes natural language parsing, natural language generation, machine translation and machine learning.

Meishan Zhang is a fifth-year PHD candidate at Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, China (HIT-SCIR). His research interest includes Chinese morphological and syntactic parsing, semantic representation and parsing, joint modeling and machine learning.

Ting Liu is a professor at HIT-SCIR. His research interest includes social computing, information retrieval and natural language processing.

to the lines of research oriented to semantic SMT, including a brief introduction of semantics, lexical and shallow semantics tailored to SMT, semantic representations in SMT, semantically motivated evaluation as well as advanced topics on deep semantic learning for SMT. The third part is dedicated to recent work on SMT with discourse, including a brief review on discourse studies from linguistics and computational viewpoints, discourse research from monolingual to multilingual, discourse-based SMT and a few advanced topics.

The tutorial is targeted for researchers in the SMT, semantics and discourse communities. In particular, the expected audience comes from two groups: 1) Researchers and students in the SMT community who want to design cutting-edge models and algorithms for semantic SMT with various semantic knowledge and representations, and who would like to advance SMT from sentence-by-sentence translation to document-level translation with discourse information; 2) Researchers and students from the semantics and discourse community who are interested in developing models and methods and adapting them to SMT.

3

Main Conference: Monday, June 23

Overview

Parallel Session 1 Overview - Monday, June 23, 2014

	Track A	Track B	Track C	Track D	Track E
	Discourse, Dialogue, Coreference and Pragmatics	*Semantics I*	*Machine Translation I*	*Syntax, Parsing, and Tagging I*	*NLP for the Web and Social Media I*
	Grand V	Grand VI	Harborside A/B	Harborside C	Harborside D/E
10:10	Representation Learning for Text-level Discourse Parsing *Ji and Eisenstein*	Multilingual Models for Compositional Distributed Semantics *Hermann and Blunsom*	Lattice Desegmentation for Statistical Machine Translation *Salameh, Cherry, and Kondrak*	[TACL] A Tabular Method for Dynamic Oracles in Transition-Based Parsing *Goldberg, Sartorio, and Satta*	Unsupervised Solution Post Identification from Discussion Forums *P and Visweswariah*
10:35	Text-level Discourse Dependency Parsing *Li, Wang, Cao, and Li*	Simple Negation Scope Resolution through Deep Parsing: A Semantic Solution to a Semantic Problem *Packard, Bender, Read, Oepen, and Dridan*	Bilingually-constrained Phrase Embeddings for Machine Translation *Zhang, Liu, Li, Zhou, and Zong*	[TACL] Joint Incremental Disfluency Detection and Dependency Parsin *Honnibal and Johnson*	Weakly Supervised User Profile Extraction from Twitter *Li, Ritter, and Hovy*
11:00	Discovering Latent Structure in Task-Oriented Dialogues *Zhai and Williams*	Logical Inference on Dependency-based Compositional Semantics *Tian, Miyao, and Matsuzaki*	Learning New Semi-Supervised Deep Auto-encoder Features for Statistical Machine Translation *Lu, Chen, and Xu*	Tagging The Web: Building A Robust Web Tagger with Neural Network *Ma, Zhang, and Zhu*	The effect of wording on message propagation: Topic- and author-controlled natural experiments on Twitter *Tan, Lee, and Pang*
11:25	Learning Structured Perceptrons for Coreference Resolution with Latent Antecedents and Non-local Features *Björkelund and Kuhn*	A practical and linguistically-motivated approach to compositional distributional semantics *Paperno, Pham, and Baroni*	Learning Topic Representation for SMT with Neural Networks *Cui, Zhang, Liu, Chen, Li, Zhou, and Yang*	[TACL] A Crossing-Sensitive Third-Order Factorization for Dependency Parsing *Pitler*	Inferring User Political Preferences from Streaming Communications *Volkova, Coppersmith, and Van Durme*

Parallel Session 1

Session 1A: Discourse, Dialogue, Coreference and Pragmatics
Grand V   *Chair: Micha Elsner*

Representation Learning for Text-level Discourse Parsing
Yangfeng Ji and Jacob Eisenstein   10:10–10:35
Text-level discourse parsing is notoriously difficult, as distinctions between discourse relations require subtle semantic judgments that are not easily captured using standard features. In this paper, we present a representation learning approach, in which we transform surface features into a latent space that facilitates RST discourse parsing. By combining the machinery of large-margin transition-based structured prediction with representation learning, our method jointly learns to parse discourse while at the same time learning a discourse-driven projection of surface features. The resulting shift-reduce discourse parser obtains substantial improvements over the previous state-of-the-art in predicting relations and nuclearity on the RST Treebank.

Text-level Discourse Dependency Parsing
Sujian Li, Liang Wang, Ziqiang Cao, and Wenjie Li   10:35–11:00
Previous researches on Text-level discourse parsing mainly made use of constituency structure to parse the whole document into one discourse tree. In this paper, we present the limitations of constituency based discourse parsing and first propose to use dependency structure to directly represent the relations between elementary discourse units (EDUs). The state-of-the-art dependency parsing techniques, the Eisner algorithm and maximum spanning tree (MST) algorithm, are adopted to parse an optimal discourse dependency tree based on the arc-factored model and the large-margin learning techniques. Experiments show that our discourse dependency parsers achieve a competitive performance on text-level discourse parsing.

Discovering Latent Structure in Task-Oriented Dialogues
Ke Zhai and Jason D Williams   11:00–11:25
A key challenge for computational models of conversation is to discover latent structure in dialogue. This is particularly true for task-oriented dialogues, where the latent structure provides a basis for analysing, evaluating, and building conversational systems. In this paper, we propose three new unsupervised dialogue models to discover latent structures in task-oriented dialogues. Our methods synthesize hidden Markov models (for underlying state) and topic models (to connect words to states). We apply our models to two corpora of real, non-trivial dialogues: human-computer spoken dialogues in the bus timetable domain, and human-human text-based web chats from a live technical support service. Qualitatively, we show that our models extract meaningful state representations and dialogue structures consistent with human annotations. Quantitatively, we evaluate using log likelihood and an ordering task on a held-out test set. We show that our models advance the state-of-the-art on both metrics, on both datasets.

Learning Structured Perceptrons for Coreference Resolution with Latent Antecedents and Non-local Features
Anders Björkelund and Jonas Kuhn   11:25–11:50
We investigate different ways of learning structured perceptron models for coreference resolution when using non-local features and beam search. Our experimental results indicate that standard techniques such as early updates or Learning as Search Optimization (LaSO) perform worse than a greedy baseline that only uses local features. By modifying LaSO to delay updates until the end of each instance we obtain significant improvements over the baseline. Our model obtains the best results to date on recent shared task data for Arabic, Chinese, and English.

Session 1B: Semantics I
Grand VI   *Chair: Mirella Lapata*

Multilingual Models for Compositional Distributed Semantics
Karl Moritz Hermann and Phil Blunsom   10:10–10:35
We present a novel technique for learning semantic representations, which extends the distributional hypothesis to multilingual data and joint-space embeddings. Our models leverage parallel data and learn to strongly align the embeddings of semantically equivalent sentences, while maintaining sufficient distance between those of dissimilar sentences. The models do not rely on word alignments or any syntactic information and are successfully applied to a number of diverse languages. We extend our approach to learn semantic representations at the document level, too. We evaluate these models on two cross-lingual document classification tasks, outperforming the prior state of the art. Through qualitative analysis and the study of pivoting effects we demonstrate that our representations are semantically plausible and can capture semantic relationships across languages without parallel data.

Simple Negation Scope Resolution through Deep Parsing: A Semantic Solution to a Semantic Problem
Woodley Packard, Emily M. Bender, Jonathon Read, Stephan Oepen, and Rebecca Dridan 10:35–11:00
In this work, we revisit Shared Task 1 from the 2012 *SEM Conference: the automated analysis of negation. Unlike the vast majority of participating systems in 2012, our approach works over explicit and formal representations of propositional semantics, i.e. derives the notion of negation scope assumed in this task from the structure of logical-form meaning representations. We relate the task-specific interpretation of (negation) scope to the concept of (quantifier and operator) scope in mainstream underspecified semantics. With reference to an explicit encoding of semantic predicate-argument structure, we can operationalize the annotation decisions made for the 2012 *SEM task, and demonstrate how a comparatively simple system for negation scope resolution can be built from an off-the-shelf deep parsing system. In a system combination setting, our approach improves over the best published results on this task to date.

Logical Inference on Dependency-based Compositional Semantics
Ran Tian, Yusuke Miyao, and Takuya Matsuzaki   11:00–11:25
Dependency-based Compositional Semantics (DCS) is a framework of natural language semantics with easy-to-process structures as well as strict semantics. In this paper, we equip the DCS framework with logical inference, by defining abstract denotations as an abstraction of the computing process of denotations in original DCS. An inference engine is built to achieve inference on abstract denotations. Furthermore, we propose a way to generate on-the-fly knowledge in logical inference, by combining our framework with the idea of tree transformation. Experiments on FraCaS and PASCAL RTE datasets show promising results.

A practical and linguistically-motivated approach to compositional distributional semantics
Denis Paperno, Nghia The Pham, and Marco Baroni   11:25–11:50
Distributional semantic methods to approximate word meaning with context vectors have been very successful empirically, and the last years have seen a surge of interest in their compositional extension to phrases and sentences. We present here a new model that, like those of Coecke et al. (2010) and Baroni and Zamparelli (2010), closely mimics the standard Montagovian semantic treatment of composition in distributional terms. However, our approach avoids a number of issues that have prevented the application of the earlier linguistically-motivated models to full-fledged, real-life sentences. We test the model on a variety of empirical tasks, showing that it consistently outperforms a set of competitive rivals.

Session 1C: Machine Translation I

Harborside A/B   *Chair: Taro Watanabe*

Lattice Desegmentation for Statistical Machine Translation

Mohammad Salameh, Colin Cherry, and Grzegorz Kondrak   10:10–10:35

Morphological segmentation is an effective sparsity reduction strategy for statistical machine translation (SMT) involving morphologically complex languages. When translating into a segmented language, an extra step is required to desegment the output; previous studies have desegmented the 1-best output from the de- coder. In this paper, we expand our translation options by desegmenting n-best lists or lattices. Our novel lattice desegmentation algorithm gives us access to both segmented and desegmented views of the target language for a large subspace of possible translation outputs. This allows us to include features related to the desegmentation process, as well as an unsegmented language model (LM), to help select the best translation. We investigate this technique in the context of English-to-Arabic and English-to-Finnish translation, showing significant improvements over desegmentation of 1-best decoder outputs.

Bilingually-constrained Phrase Embeddings for Machine Translation

Jiajun Zhang, Shujie Liu, Mu Li, Ming Zhou, and Chengqing Zong   10:35–11:00

We propose Bilingually-constrained Recursive Auto-encoders (BRAE) to learn semantic phrase embeddings (compact vector representations for phrases), which can distinguish the phrases with different semantic meanings. The BRAE is trained in a way that minimizes the semantic distance of translation equivalents and maximizes the semantic distance of non-translation pairs simultaneously. After training, the model learns how to embed each phrase semantically in two languages and also learns how to transform semantic embedding space in one language to the other. We evaluate our proposed method on two end-to-end SMT tasks (phrase table pruning and decoding with phrasal semantic similarities) which need to measure semantic similarity between a source phrase and its translation candidates. Extensive experiments show that the BRAE is remarkably effective in these two tasks.

Learning New Semi-Supervised Deep Auto-encoder Features for Statistical Machine Translation

Shixiang Lu, Zhenbiao Chen, and Bo Xu   11:00–11:25

In this paper, instead of designing new features based on intuition, linguistic knowledge and domain, we learn some new and effective features using the deep auto-encoder (DAE) paradigm for phrase-based translation model. Using the unsupervised pre-trained deep belief net (DBN) to initialize DAE's parameters and using the input original phrase features as a teacher for semi-supervised fine-tuning, we learn new semi-supervised DAE features, which are more effective and stable than the unsupervised DBN features. Moreover, to learn high dimensional feature representation, we introduce a natural horizontal composition of more DAEs for large hidden layers feature learning. On two Chinese-English tasks, our semi-supervised DAE features obtain statistically significant improvements of 1.34/2.45 (IWSLT) and 0.82/1.52 (NIST) BLEU points over the unsupervised DBN features and the baseline features, respectively.

Learning Topic Representation for SMT with Neural Networks

Lei Cui, Dongdong Zhang, Shujie Liu, Qiming Chen, Mu Li, Ming Zhou, and Muyun Yang 11:25–11:50

Statistical Machine Translation (SMT) usually utilizes contextual information to disambiguate translation candidates. However, it is often limited to contexts within sentence boundaries, hence broader topical information cannot be leveraged. In this paper, we propose a novel approach to learning topic representation for parallel data using a neural network architecture, where abundant topical contexts are embedded via topic relevant monolingual data. By associating each translation rule with the topic representation, topic relevant rules are selected according to the distributional similarity with the source text during SMT decoding. Experimental results show that our method significantly improves translation accuracy in the NIST Chinese-to-English translation task compared to a state-of-the-art baseline.

Session 1D: Syntax, Parsing, and Tagging I
Harborside C   *Chair: Ting Liu*

[TACL] Training Deterministic Parsers with Non-Deterministic Oracles
Yoav Goldberg and Joakim Nivre   10:10–10:35

Greedy transition-based parsers are very fast but tend to suffer from error propagation. This problem is aggravated by the fact that they are normally trained using oracles that are deterministic and incomplete in the sense that they assume a unique canonical path through the transition system and are only valid as long as the parser does not stray from this path. In this paper, we give a general characterization of oracles that are nondeterministic and complete, present a method for deriving such oracles for transition systems that satisfy a property we call arc decomposition, and instantiate this method for three well-known transition systems from the literature. We say that these oracles are dynamic, because they allow us to dynamically explore alternative and nonoptimal paths during training — in contrast to oracles that statically assume a unique optimal path. Experimental evaluation on a wide range of data sets clearly shows that using dynamic oracles to train greedy parsers gives substantial improvements in accuracy. Moreover, this improvement comes at no cost in terms of efficiency, unlike other techniques like beam search.

[TACL] Joint Incremental Disfluency Detection and Dependency Parsin
Matthew Honnibal and Mark Johnson   10:35–11:00

We present an incremental dependency parsing model that jointly performsdisfluency detection. The model handles speech repairs using a novel non-monotonictransition system, and includes several novel classes of features. Together, these additions improve parse accuracy by 1%. The model runs in expected linear time, and processes over 700 sentences a second. For comparison, we evaluated two pipeline systems, using state-of-the-art disfluencydetectors. The joint model performed better on both tasks,with a parse accuracy of 90.5% and 84.0% accuracy at disfluency detection.

Tagging The Web: Building A Robust Web Tagger with Neural Network
Ji Ma, Yue Zhang, and Jingbo Zhu   11:00–11:25

In this paper, we address the problem of web-domain POS tagging using a two- phase approach. The first phase learns representations that capture regularities underlying web text. The representation is integrated as features into a neural network that serves as a scorer for an easy-first POS tagger. Parameters of the neural network are trained using guided learning in the second phase. Experiment on the SANCL 2012 shared task show that our approach achieves 93.27% average tagging accuracy, which is the best accuracy reported so far on this data set, higher than those given by ensembled syntactic parsers.

[TACL] A Crossing-Sensitive Third-Order Factorization for Dependency Parsing
Emily Pitler   11:25–11:50

Features over grandparents and siblings have greatly improved the accuracy of projective parsers, but have so far required large increases in complexity when applied to arborescences or mildly non-projective models. We introduce a "crossing-sensitive" generalization of the third-order factorization of Koo and Collins (2010) that trades off complexity in the model structure (i.e. scoring with features over pairs and triples of edges) with complexity in the output structure (i.e. producing crossing edges). When applied to a projective tree, the crossing-sensitive factorization exactly simplifies to Koo and Collins' Grand-Sibling model. Under this model, the optimal 1-Endpoint-Crossing tree (Pitler et al., 2013) can be found in $O(n^4)$ time, matching the asymptotic run-time of both the third- order projective parser and the edge-factored 1-Endpoint-Crossing parser. The crossing- sensitive third-order parser is significantly more accurate than the third-order projective parser under many experimental settings and significantly less accurate on none. Besides the asymptotic guarantees, the cost of running the crossing-sensitive third-order parser is low in practice, running at 0.37-0.47 times the speed of the third-order projective parser.

Session 1E: NLP for the Web and Social Media I

Harborside D/E   *Chair: Miles Osborne*

Unsupervised Solution Post Identification from Discussion Forums
Deepak P and Karthik Visweswariah   10:10–10:35

Discussion forums have evolved into a dependable source of knowledge to solve common problems. However, only a minority of the posts in discussion forums are solution posts. Identifying solution posts from discussion forums, hence, is an important research problem. In this paper, we present a technique for unsupervised solution post identification leveraging a so far unexplored textual feature, that of lexical correlations between problems and solutions. We use translation models and language models to exploit lexical correlations and solution post character respectively. Our technique is designed to not rely much on structural features such as post metadata since such features are often not uniformly available across forums. Our clustering-based iterative solution identification approach based on the EM-formulation performs favorably in an empirical evaluation, beating the only unsupervised solution identification technique from literature by a very large margin. We also show that our unsupervised technique is competitive against methods that require supervision, outperforming one such technique comfortably.

Weakly Supervised User Profile Extraction from Twitter
Jiwei Li, Alan Ritter, and Eduard Hovy   10:35–11:00

While user attribute extraction on social media has received considerable research attention, existing approaches, mostly supervised, encounter great difficulty in obtaining gold standards, thus usually narrowed to trivial domains where ground truth is easy to identify (e.g., gender). In this paper, we present a weakly-supervised approach to for user profile extraction from Twitter. Users' profiles from social media websites such as Facebook or Google Plus are used as a distant source of supervision for extraction of their attributes from user-generated text. In addition to traditional linguistic features used in distant supervision for information extraction, our approach also takes into account network information, a unique opportunity offered by social media. We test our algorithm on three attribute domains: spouse, education and job; experimental results demonstrate our approach is able to make accurate predictions for users' attributes based on their tweets.

The effect of wording on message propagation: Topic- and author-controlled natural experiments on Twitter
Chenhao Tan, Lillian Lee, and Bo Pang   11:00–11:25

Consider a person trying to spread an important message on a social network. He/she can spend hours trying to craft the message. Does it actually matter? While there has been extensive prior work looking into predicting popularity of social-media content, the effect of wording per se has rarely been studied since it is of- ten confounded with the popularity of the author and the topic. To control for these confounding factors, we take advantage of the surprising fact that there are many pairs of tweets containing the same url and written by the same user but employing different wording. Given such pairs, we ask: which version attracts more retweets? This turns out to be a more difficult task than predicting popular topics. Still, humans can answer this question better than chance (but far from perfectly), and the computational methods we develop can do better than an average human as well as a strong competing method trained on non-controlled data.

Inferring User Political Preferences from Streaming Communications
Svitlana Volkova, Glen Coppersmith, and Benjamin Van Durme   11:25–11:50

Existing models for social media personal analytics assume access to thousands of messages per user, even though most users author content only sporadically over time. Given this sparsity, we: (i) leverage content from the local neighborhood of a user; (ii) evaluate batch models as a function of size and the amount of messages in various types of neighborhoods; and (iii) estimate the amount of time and tweets required for a dynamic model to predict user preferences. We show that even when limited or no self-authored data is available, language from friend, retweet and user mention communications provide sufficient evidence for prediction. When updating models over time based on Twitter, we find that political preference can be often be predicted using roughly 100 tweets, depending on the context of user selection, where this could mean hours, or weeks, based on the author's tweeting frequency.

Parallel Session 2 Overview - Monday, June 23, 2014

	Track A	Track B	Track C	Track D	Track E
	Syntax, Parsing and Tagging II	*Semantics II*	*Word Segmentation and POS Tagging*	*SRW Thesis Proposals*	*Sentiment Analysis I*
	Grand V	Grand VI	Harborside A/B	Harborside C	Harborside D/E
1:30	Steps to Excellence: Simple Inference with Refined Scoring of Dependency Trees *Zhang, Lei, Barzilay, Jaakkola, and Globerson*	Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors *Baroni, Dinu, and Kruszewski*	[TACL] Exploring the Role of Stress in Bayesian Word Segmentation using Adaptor Grammars *Börschinger and Johnson*	Bayesian Kernel Methods for Natural Language Processing *Beck*	An Empirical Study on the Effect of Negation Words on Sentiment *Zhu, Guo, Mohammad, and Kiritchenko*
1:45	Sparser, Better, Faster GPU Parsing *Hall, Berg-Kirkpatrick, and Klein*	Metaphor Detection with Cross-Lingual Model Transfer *Tsvetkov, Boytsov, Gershman, Nyberg, and Dyer*	Modelling function words improves unsupervised word segmentation *Johnson, Christophe, Dupoux, and Demuth*	Extracting Temporal and Causal Relations between Events *Mirza*	Extracting Opinion Targets and Opinion Words from Online Reviews with Graph Co-ranking *Liu, Xu, and Zhao*
2:10	Shift-Reduce CCG Parsing with a Dependency Model *Xu, Clark, and Zhang*	Learning Word Sense Distributions, Detecting Unattested Senses and Identifying Novel Senses Using Topic Models *Lau, Cook, McCarthy, Gella, and Baldwin*	[TACL] FLORS: Fast and Simple Domain Adaptation for Part-of-Speech Tagging *Schnabel and Schütze*	Towards a discourse relation-aware approach for Chinese-English machine translation *Yung*	Context-aware Learning for Sentence-level Sentiment Analysis with Posterior Regularization *Yang and Cardie*
2:35	Less Grammar, More Features *Hall, Durrett, and Klein*	Learning to Automatically Solve Algebra Word Problems *Kushman, Zettlemoyer, Barzilay, and Artzi*	Max-Margin Tensor Neural Network for Chinese Word Segmentation *Pei, Ge, and Chang*	Analyzing Positions and Topics in Political Discussions of the German Bundestag *Zirn*	Product Feature Mining: Semantic Clues versus Syntactic Constituents *Xu, Liu, Lai, and Zhao*

Parallel Session 2

Session 2A: Syntax, Parsing and Tagging II
Grand V   *Chair: Julia Hockenmaier*

Steps to Excellence: Simple Inference with Refined Scoring of Dependency Trees
Yuan Zhang, Tao Lei, Regina Barzilay, Tommi Jaakkola, and Amir Globerson   1:30–1:45

Much of the recent work on dependency parsing has been focused on solving inherent combinatorial problems associated with rich scoring functions. In contrast, we demonstrate that highly expressive scoring functions can be used with substantially simpler inference procedures. Specifically, we introduce a sampling-based parser that can easily handle arbitrary global features. Inspired by SampleRank, we learn to take guided stochastic steps towards a high scoring parse. We introduce two samplers for traversing the space of trees, Gibbs and Metropolis-Hastings with Random Walk. The model outperforms state-of-the-art results when evaluated on 14 languages of non-projective CoNLL datasets. Our sampling-based approach naturally extends to joint prediction scenarios, such as joint parsing and POS correction. The resulting method outperforms the best reported results on the CATiB dataset, approaching performance of parsing with gold tags.

Sparser, Better, Faster GPU Parsing
David Hall, Taylor Berg-Kirkpatrick, and Dan Klein   1:45–2:10

Due to their origin in computer graphics, graphics processing units (GPUs) are highly optimized for dense problems, where the exact same operation is applied repeatedly to all data points. Natural language processing algorithms, on the other hand, are traditionally constructed in ways that exploit structural sparsity. Recently, Canny et al. (2013) presented an approach to GPU parsing that sacrifices traditional sparsity in exchange for raw computational power, obtaining a system that can compute Viterbi parses for a high-quality grammar at about 164 sentences per second on a mid-range GPU. In this work, we reintroduce sparsity to GPU parsing by adapting a coarse-to-fine pruning approach to the constraints of a GPU. The resulting system is capable of computing over 404 Viterbi parses per second—more than a 2x speedup—on the same hardware. Moreover, our approach allows us to efficiently implement less GPU-friendly minimum Bayes risk inference, improving throughput for this more accurate algorithm from only 32 sentences per second unpruned to over 190 sentences per second using pruning—nearly a 6x speedup.

Shift-Reduce CCG Parsing with a Dependency Model
Wenduan Xu, Stephen Clark, and Yue Zhang   2:10–2:35

This paper presents the first dependency model for a shift-reduce CCG parser. Modelling dependencies is desirable for a number of reasons, including handling the "spurious" ambiguity of CCG; fitting well with the theory of CCG; and optimizing for structures which are evaluated at test time. We develop a novel training technique using a dependency oracle, in which all derivations are hidden. A challenge arises from the fact that the oracle needs to keep track of exponentially many gold-standard derivations, which is solved by integrating a packed parse forest with the beam-search decoder. Standard CCGBank tests show the model achieves up to 1.05 labeled F-score improvements over three existing, competitive CCG parsing models.

Less Grammar, More Features
David Hall, Greg Durrett, and Dan Klein   2:35–3:00

We present a parser that relies primarily on extracting information directly from surface spans rather than on propagating information through enriched grammar structure. For example, instead of creating separate grammar symbols to mark the definiteness of an NP, our parser might instead capture the same information from the first word of the NP. Moving context out of the grammar and onto surface features can greatly simplify the structural component of the parser: because so many deep syntactic cues have surface reflexes, our system can still parse accurately with context-free backbones as minimal as X-bar grammars. Keeping the structural backbone simple and moving features to the surface also allows easy adaptation to new languages and even to new tasks. On the SPMRL 2013 multilingual constituency parsing shared task (Seddah et al., 2013), our system outperforms the top single parser system of Björkelund et al. (2013) on a range of languages. In addition, despite being designed for syntactic analysis, our system also achieves state-of-the-art numbers on the structural sentiment task of Socher et al. (2013). Finally, we show that, in both syntactic parsing and sentiment analysis, many broad linguistic trends can be captured via surface features.

Session 2B: Semantics II
Grand VI   *Chair: Raymond Mooney*

Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors
Marco Baroni, Georgiana Dinu, and Germán Kruszewski   1:30–1:45

Context-predicting models (more commonly known as embeddings or neural language models) are the new kids on the distributional semantics block. Despite the buzz surrounding these models, the literature is still lacking a systematic comparison of the predictive models with classic, count-vector-based distributional semantic approaches. In this paper, we perform such an extensive evaluation, on a wide range of lexical semantics tasks and across many parameter settings. The results, to our own surprise, show that the buzz is fully justified, as the context-predicting models obtain a thorough and resounding victory against their count-based counterparts.

Metaphor Detection with Cross-Lingual Model Transfer
Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer   1:45–2:10

We show that it is possible to reliably discriminate whether a syntactic construction is meant literally or metaphorically using lexical semantic features of the words that participate in the construction. Our model is constructed using English resources, and we obtain state-of-the-art performance relative to previous work in this language. Using a model transfer approach by pivoting through a bilingual dictionary, we show our model can identify metaphoric expressions in other languages. We provide results on three new test sets in Spanish, Farsi, and Russian. The results support the hypothesis that metaphors are conceptual, rather than lexical, in nature.

Learning Word Sense Distributions, Detecting Unattested Senses and Identifying Novel Senses Using Topic Models
Jey Han Lau, Paul Cook, Diana McCarthy, Spandana Gella, and Timothy Baldwin   2:10–2:35

Unsupervised word sense disambiguation (WSD) methods are an attractive approach to all-words WSD due to their non-reliance on expensive annotated data. Estimates of sense frequency have been shown to be very useful for WSD due to the skewed nature of word sense distributions. This paper presents a fully unsupervised topic modelling-based approach to sense frequency estimation, which is highly portable to different corpora and sense inventories, in being applicable to any part of speech, and not requiring a hierarchical sense inventory, parsing or parallel text. We demonstrate the effectiveness of the method over the tasks of predominant sense learning and sense distribution acquisition, and also the novel tasks of detecting senses which aren't attested in the corpus, and identifying novel senses in the corpus which aren't captured in the sense inventory.

Learning to Automatically Solve Algebra Word Problems
Nate Kushman, Luke Zettlemoyer, Regina Barzilay, and Yoav Artzi   2:35–3:00

We present an approach for automatically learning to solve algebra word problems given only text. Successfully solving these problems requires handling complex semantic relations which cross sentences, such as determining that the distinct entities mentioned in various sentences are all a part of the same set. Our algorithm first determines the relation between entities in the text, and then translates these into a system of equations. The algorithm is trained in a weak supervision setting, where we are only given the text and the resulting numerical answers, without information on the entity relations or the underlying equations. Despite this relatively weak supervision, we are able to successfully generate the correct answer set for 77% of the problems on a dataset of actual grade-school algebra problems, compared to the 26% produced by a baseline which is agnostic to semantic relations.

Session 2C: Word Segmentation and POS Tagging

Harborside A/B   Chair: Grzegorz Kondrak

[TACL] Exploring the Role of Stress in Bayesian Word Segmentation using Adaptor Grammars

Benjamin Börschinger and Mark Johnson   1:20–1:45

Stress has long been established as a major cue in word segmentation for English infants. We show that enabling a current state-of-the-artBayesian word segmentation model to take advantage of stress cuesnoticeably improves its performance. We find that the improvementsrange from 10 to 4%, depending on both the use of phonotactic cuesand, to a lesser extent, the amount of evidence available to the learner.We also find that in particular early on, stress cues are much moreuseful for our model than phonotactic cues by themselves, consistent with the finding that children do seem to use stress cues before they use phonotactic cues. Finally, we study how the model's knowledge about stress patterns evolves over time. We not only find that our model correctly acquires the most frequent patterns relatively quickly but also that the Unique Stress Constraint that is at the heart of a previously proposed model does not need to be built in but can be acquired jointly with word segmentation.

Modelling function words improves unsupervised word segmentation

Mark Johnson, Anne Christophe, Emmanuel Dupoux, and Katherine Demuth   1:45–2:10

Inspired by experimental psychological findings suggesting that function words play a special role in word learning, we make a simple modification to an Adaptor Grammar based Bayesian word segmentation model to allow it to learn sequences of monosyllabic "function words" at the beginnings and endings of collocations of (possibly multi-syllabic) words. This modification improves unsupervised word segmentation on the standard Bernstein-Ratner (1987) corpus of child-directed English by more than 4% token f-score compared to a model identical except that it does not special-case "function words", setting a new state-of-the-art of 92.4% token f-score. Our function word model assumes that function words appear at the left periphery, and while this is true of languages such as English, it is not true universally. We show that a learner can use Bayesian model selection to determine the location of function words in their language, even though the input to the model only consists of unsegmented sequences of phones. Thus our computational models support the hypothesis that function words play a special role in word learning.

[TACL] FLORS: Fast and Simple Domain Adaptation for Part-of-Speech Tagging

Tobias Schnabel and Hinrich Schütze   2:10–2:35

We present FLORS, a new part-of-speech tagger for domain adaptation. FLORS uses robust representations that work especially well for unknown words and for known words with unseen tags. FLORS is simpler and faster than previous domain adaptation methods, yet it has significantly better accuracy than several baselines.

Max-Margin Tensor Neural Network for Chinese Word Segmentation

Wenzhe Pei, Tao Ge, and Baobao Chang   2:35–3:00

Recently, neural network models for natural language processing tasks have been increasingly focused on for their ability to alleviate the burden of manual feature engineering. In this paper, we propose a novel neural network model for Chinese word segmentation called Max-Margin Tensor Neural Network (MMTNN). By exploiting tag embedding and tensor-based transformation, MMTNN has the ability to model complicated interactions between tags and context characters. Furthermore, a new tensor factorization approach is proposed to speed up the model and avoid overfitting. Experiments on the benchmark dataset show that our model achieves better performances than previous neural network models and that our model can achieve a competitive performance with minimal feature engineering. Despite Chinese word segmentation being a specific case, MMTNN can be easily generalized and applied to other sequence labeling tasks.

Session 2D: SRW Thesis Proposals

Harborside C   *Chair: Svitlana Volkova*

Bayesian Kernel Methods for Natural Language Processing
Daniel Beck   1:20–1:45
Kernel methods are heavily used in Natural Language Processing (NLP). Frequentist approaches like Support Vector Machines are the state-of-the-art in many tasks. However, these approaches lack efficient procedures for model selection, which hinders the usage of more advanced kernels. In this work, we propose the use of a Bayesian approach for kernel methods, Gaussian Processes, which allow easy model fitting even for complex kernel combinations. Our goal is to employ this approach to improve a number of regression and classification tasks in NLP.

Extracting Temporal and Causal Relations between Events
Paramita Mirza   1:45–2:10
A notably challenging problem related to event processing is recognizing the relations holding between events in a text, in particular temporal and causal relations. While there has been some research on temporal relations, the aspect of causality between events from a Natural Language Processing (NLP) perspective has hardly been touched. We propose an annotation scheme to cover different types of causality between events, techniques for extracting such relations and an investigation into the connection between temporal and causal relations. In this thesis work we aim to focus especially on the latter, because causality is presumed to have a temporal constraint. We conjecture that injecting this presumption may be beneficial for the recognition of both temporal and causal relations.

Towards a discourse relation-aware approach for Chinese-English machine translation
Frances Yung   2:10–2:35
Translation of discourse relations is one of the recent efforts of incorporating discourse information to statistical machine translation (SMT). While existing works focus on disambiguation of ambiguous discourse connectives, or transformation of discourse trees, only explicit discourse relations are tackled. A greater challenge exists in machine translation of Chinese, since implicit discourse relations are abundant and occur both inside and outside a sentence. This thesis proposal describes ongoing work on bilingual discourse annotation and plans towards incorporating discourse relation knowledge to an SMT system for Chinese and English with consideration of implicit discourse relations. The final goal is a discourse-unit-based translation model that disregards the traditional assumption of sentence-to-sentence translation.

Analyzing Positions and Topics in Political Discussions of the German Bundestag
Cäcilia Zirn   2:35–3:00
We present ongoing doctoral work on automatically understanding the positions of politicians with respect to those of the party they belong to. To this end, we use textual data, namely transcriptions of political speeches from meetings of the German Bundestag, and party manifestos, in order to automatically acquire the positions of political actors and parties, respectively. We discuss a variety of possible supervised and unsupervised approaches to determine the topics of interest and compare positions, and propose to explore an approach based on topic modeling techniques for these tasks.

Session 2E: Sentiment Analysis I   Harborside D/E   *Chair: Myle Ott*

An Empirical Study on the Effect of Negation Words on Sentiment
Xiaodan Zhu, Hongyu Guo, Saif Mohammad, and Svetlana Kiritchenko   1:30–1:45

Negation words, such as no and not, play a fundamental role in modifying sentiment of textual expressions. We will refer to a negation word as the negator and the text span within the scope of the negator as the argument. Commonly used heuristics to estimate the sentiment of negated expressions rely simply on the sentiment of argument (and not on the negator or the argument itself). We use a sentiment treebank to show that these existing heuristics are poor estimators of sentiment. We then modify these heuristics to be dependent on the negators and show that this improves prediction. Next, we evaluate a recently proposed composition model (Socher et al., 2013) that relies on both the negator and the argument. This model learns the syntax and semantics of the negator's argument with a recursive neural network. We show that this approach performs better than those mentioned above. Finally, we explicitly incorporate the prior sentiment of the argument and observe that this information can help further reduce fitting errors.

Extracting Opinion Targets and Opinion Words from Online Reviews with Graph Co-ranking
Kang Liu, Liheng Xu, and Jun Zhao   1:45–2:10

Extracting opinion targets and opinion words from online reviews are two fundamental tasks in opinion mining. This paper proposes a novel approach to collectively extract them with graph co-ranking. First, compared to previous methods which solely employed opinion relations among words, our method constructs a heterogeneous graph to model two types of relations, including semantic relations and opinion relations. Next, a co-ranking algorithm is proposed to estimate the confidence of each candidate, and the candidates with higher confidence will be extracted as opinion targets/words. In this way, different relations make cooperative effects on confidence estimation of each candidate. Moreover, word preference is captured and incorporated into our co-ranking algorithm. In this way, our co-ranking is personalized and each candidate confidence is only determined by their preferred collocations, which helps to improve the extraction precision. The experimental results on three data sets with different sizes and languages show that our approach achieves better performance than state-of-the-art methods.

Context-aware Learning for Sentence-level Sentiment Analysis with Posterior Regularization
Bishan Yang and Claire Cardie   2:10–2:35

This paper proposes a novel context-aware method for analyzing sentiment at the level of individual sentences. Most existing machine learning approaches suffer from limitations in the modeling of complex linguistic structures across sentences and often fail to capture non-local contextual cues that are important for sentiment interpretation. In contrast, our approach allows structured modeling of sentiment while taking into account both local and global contextual information. Specifically, we encode intuitive lexical and discourse knowledge as expressive constraints and integrate them into the learning of conditional random field models via posterior regularization. The context-aware constraints provide additional power to the CRF model and can guide semi-supervised learning when labeled data is limited. Experiments on standard product review datasets show that our method outperforms the state-of-the-art methods in both the supervised and semi-supervised settings.

Product Feature Mining: Semantic Clues versus Syntactic Constituents
Liheng Xu, Kang Liu, Siwei Lai, and Jun Zhao   2:35–3:00

Product feature mining is a key subtask in fine-grained opinion mining. Previous works often use syntax constituents in this task. However, syntax-based methods can only use discrete contextual information, which may suffer from data sparsity. This paper proposes a novel product feature mining method which leverages lexical and contextual semantic clues. Lexical semantic clue verifies whether a candidate term is related to the target product, and contextual semantic clue serves as a soft pattern miner to find candidates, which exploits semantics of each word in context so as to alleviate the data sparsity problem. We build a semantic similarity graph to encode lexical semantic clue, and employ a convolutional neural model to capture contextual semantic clue. Then Label Propagation is applied to combine both semantic clues. Experimental results show that our semantics-based method significantly outperforms conventional syntax-based approaches, which not only mines product features more accurately, but also extracts more infrequent product features.

Parallel Session 3 Overview - Monday, June 23, 2014

	Track A	Track B	Track C	Track D	Track E
	Topic Modeling	*Information Extraction I*	*Generation*	*Syntax, Parsing, and Tagging III*	*Language Resources and Evaluation I*
	Grand V	Grand VI	Harborside A/B	Harborside C	Harborside D/E
3:30	Aspect Extraction with Automated Prior Knowledge Learning *Chen, Mukherjee, and Liu*	Collective Tweet Wikification based on Semi-supervised Graph Regularization *Huang, Cao, Huang, Ji, and Lin*	That's Not What I Meant! Using Parsers to Avoid Structural Ambiguities in Generated Text *Duan and White*	[TACL] Training Deterministic Parsers with Non-Deterministic Oracles *Goldberg and Nivre*	A Robust Approach to Aligning Heterogeneous Lexical Resources *Pilehvar and Navigli*
3:55	Anchors Regularized: Adding Robustness and Extensibility to Scalable Topic-Modeling Algorithms *Nguyen, Hu, and Boyd-Graber*	Zero-shot Entity Extraction from Web Pages *Pasupat and Liang*	Surface Realisation from Knowledge-Bases *Gyawali and Gardent*	Grammatical Relations in Chinese: GB-Ground Extraction and Data-Driven Parsing *Sun, Du, Kou, Ding, and Wan*	Predicting the relevance of distributional semantic similarity with contextual information *Muller, Fabre, and Adam*
4:20	A Bayesian Mixed Effects Model of Literary Character *Bamman, Underwood, and Smith*	Incremental Joint Extraction of Entity Mentions and Relations *Li and Ji*	Hybrid Simplification using Deep Semantics and Machine Translation *Narayan and Gardent*	Ambiguity-aware Ensemble Training for Semi-supervised Dependency Parsing *Li, Zhang, and Chen*	[TACL] Temporal Annotation in the Clinical Domain *Styler, Bethard, Finan, Palmer, Pradhan, Groen, Erickson, Miller, Lin, Savova, and Pustejovsky*

Parallel Session 3

Session 3A: Topic Modeling
Grand V   *Chair: Chris Dyer*

Aspect Extraction with Automated Prior Knowledge Learning
Zhiyuan Chen, Arjun Mukherjee, and Bing Liu   3:30–3:55

Aspect extraction is an important task in sentiment analysis. Topic modeling is a popular method for the task. However, unsupervised topic models often generate incoherent aspects. To address the issue, several knowledge-based models have been proposed to incorporate prior knowledge provided by the user to guide modeling. In this paper, we take a major step forward and show that in the big data era, without any user input, it is possible to learn prior knowledge automatically from a large amount of review data available on the Web. Such knowledge can then be used by a topic model to discover more coherent aspects. There are two key challenges: (1) learning quality knowledge from reviews of diverse domains, and (2) making the model fault-tolerant to handle possibly wrong knowledge. A novel approach is proposed to solve these problems. Experimental results using reviews from 36 domains show that the proposed approach achieves significant improvements over state-of-the-art baselines.

Anchors Regularized: Adding Robustness and Extensibility to Scalable Topic-Modeling Algorithms
Thang Nguyen, Yuening Hu, and Jordan Boyd-Graber   3:55–4:20

Spectral methods offer scalable alternatives to Markov chain Monte Carlo and expectation maximization. However, these new methods lack the rich priors associated with probabilistic models. We examine Arora et al.'s anchor words algorithm for topic modeling and develop new, regularized algorithms that not only mathematically resemble Gaussian and Dirichlet priors but also improve the interpretability of topic models. Our new regularization approaches make these efficient algorithms more flexible; we also show that these methods can be combined with informed priors.

A Bayesian Mixed Effects Model of Literary Character
David Bamman, Ted Underwood, and Noah A. Smith   4:20–4:45

We consider the problem of automatically inferring latent character types in a collection of 15,099 English novels published between 1700 and 1899. Unlike prior work in which character types are assumed responsible for probabilistically generating all text associated with a character, we introduce a model that employs multiple effects to account for the influence of extra-linguistic information (such as author). In an empirical evaluation, we find that this method leads to improved agreement with the preregistered judgments of a literary scholar, complementing the results of alternative models.

Session 3B: Information Extraction I
Grand VI   *Chair: Mihai Surdeanu*

Collective Tweet Wikification based on Semi-supervised Graph Regularization
Hongzhao Huang, Yunbo Cao, Xiaojiang Huang, Heng Ji, and Chin-Yew Lin   3:30–3:55

Wikification for tweets aims to automatically identify each concept mention in a tweet and link it to a concept referent in a knowledge base (e.g., Wikipedia). Due to the shortness of a tweet, a collective inference model incorporating global evidence from multiple mentions and concepts is more appropriate than a non-collecitve approach which links each mention at a time. In addition, it is challenging to generate sufficient high quality labeled data for supervised models with low cost. To tackle these challenges, we propose a novel semi-supervised graph regularization model to incorporate both local and global evidence from multiple tweets through three novel fine-grained relations. In order to identify semantically-related mentions for collective inference, we detect meta path-based semantic relations through social networks. Compared to the state-of-the-art supervised model trained from 100% labeled data, our proposed approach achieves comparable performance with 31% labeled data and obtains 5% absolute F1 gain with 50% labeled data.

Zero-shot Entity Extraction from Web Pages
Panupong Pasupat and Percy Liang   3:55–4:20

In order to extract entities of a fine-grained category from semi-structured data in web pages, existing information extraction systems rely on seed examples or redundancy across multiple web pages. In this paper, we consider a new zero-shot learning task of extracting entities specified by a natural language query (in place of seeds) given only a single web page. Our approach defines a log-linear model over latent extraction predicates, which select lists of entities from the web page. The main challenge is to define features on widely varying candidate entity lists. We tackle this by abstracting list elements and using aggregate statistics to define features. Finally, we created a new dataset of diverse queries and web pages, and show that our system achieves significantly better accuracy than a natural baseline.

Incremental Joint Extraction of Entity Mentions and Relations
Qi Li and Heng Ji   4:20–4:45

We present an incremental joint frame- work to simultaneously extract entity mentions and relations using structured perceptron with efficient beam-search. A segment-based decoder based on the idea of semi-Markov chain is adopted to the new framework as opposed to traditional token-based tagging. In addition, by virtue of the inexact search, we developed a number of new and effective global features as soft constraints to capture the inter- dependency among entity mentions and relations. Experiments on Automatic Content Extraction (ACE) corpora demonstrate that our joint model significantly outperforms a strong pipelined baseline, which attains better performance than the best-reported end-to-end system.

Session 3C: Generation
Harborside A/B   *Chair: Margaret Mitchell*

That's Not What I Meant! Using Parsers to Avoid Structural Ambiguities in Generated Text
Manjuan Duan and Michael White   3:30–3:55

We investigate whether parsers can be used for self-monitoring in surface realization in order to avoid egregious errors involving "vicious" ambiguities, namely those where the intended interpretation fails to be considerably more likely than alternative ones. Using parse accuracy in a simple reranking strategy for self-monitoring, we find that with a state-of-the-art averaged perceptron realization ranking model, BLEU scores cannot be improved with any of the well-known Treebank parsers we tested, since these parsers too often make errors that human readers would be unlikely to make. However, by using an SVM ranker to combine the realizer's model score together with features from multiple parsers, including ones designed to make the ranker more robust to parsing mistakes, we show that significant increases in BLEU scores can be achieved. Moreover, via a targeted manual analysis, we demonstrate that the SVM reranker frequently manages to avoid vicious ambiguities, while its ranking errors tend to affect fluency much more often than adequacy.

Surface Realisation from Knowledge-Bases
Bikash Gyawali and Claire Gardent   3:55–4:20

We present a simple, data-driven approach to generation from knowledge bases (KB). A key feature of this approach is that grammar induction is driven by the extended domain of locality principle of TAG (Tree Adjoining Grammar); and that it takes into account both syntactic and semantic information. The resulting extracted TAG includes a unification based semantics and can be used by an existing surface realiser to generate sentences from KB data. Experimental evaluation on the KBGen data shows that our model outperforms a data-driven generate-and-rank approach based on an automatically induced probabilistic grammar; and is comparable with a handcrafted symbolic approach.

Hybrid Simplification using Deep Semantics and Machine Translation
Shashi Narayan and Claire Gardent   4:20–4:45

We present a hybrid approach to sentence simplification which combines deep semantics and monolingual machine translation to derive simple sentences from complex ones. The approach differs from previous work in two main ways. First, it is semantic based in that it takes as input a deep semantic representation rather than e.g., a sentence or a parse tree. Second, it combines a simplification model for splitting and deletion with a monolingual translation model for phrase substitution and reordering. When compared against current state of the art methods, our model yields significantly simpler output that is both grammatical and meaning preserving.

Session 3D: Syntax, Parsing, and Tagging III
Harborside C   *Chair: Yusuke Miyao*

[TACL] A Tabular Method for Dynamic Oracles in Transition-Based Parsing
Yoav Goldberg, Francesco Sartorio, and Giorgio Satta   3:30–3:55

We develop parsing oracles for two transition-based dependency parsers, including the arc-standard parser, solving a problem that was left open in (Goldberg and Nivre, 2013). We experimentally show that using these oracles in training yields superior parsing accuracies on many languages.

Grammatical Relations in Chinese: GB-Ground Extraction and Data-Driven Parsing
Weiwei Sun, Yantao Du, Xin Kou, Shuoyang Ding, and Xiaojun Wan   3:55–4:20

This paper is concerned with building linguistic resources and statistical parsers for deep grammatical relation (GR) analysis of Chinese texts. A set of linguistic rules that are grounded in the Government and Binding theory is defined to explore implicit phrase structural information and thus build high-quality GR annotations that are represented as general directed dependency graphs. The reliability of this linguistically-motivated GR extraction procedure is highlighted by manual evaluation. Based on the converted corpus, we study transition-based, data-driven models for GR parsing. We present a novel transition system which suits GR graphs better than existing systems. The key idea is to introduce a new type of transition that reorders top k elements in the memory module. Evaluation gauges how successful GR parsing for Chinese can be by applying data-driven models.

Ambiguity-aware Ensemble Training for Semi-supervised Dependency Parsing
Zhenghua Li, Min Zhang, and Wenliang Chen   4:20–4:45

This paper proposes a simple yet effective framework for semi-supervised dependency parsing at entire tree level, referred to as ambiguity-aware ensemble training. Instead of only using 1-best parse trees in previous work, our core idea is to utilize parse forest (ambiguous labelings) to combine multiple 1-best parse trees generated from diverse parsers on unlabeled data. With a conditional random field based probabilistic dependency parser, our training objective is to maximize mixed likelihood of labeled data and auto-parsed unlabeled data with ambiguous labelings. This framework offers two promising advantages. 1) ambiguity encoded in parse forests compromises noise in 1-best parse trees. During training, the parser is aware of these ambiguous structures, and has the flexibility to distribute probability mass to its preferred parse trees as long as the likelihood improves. 2) diverse syntactic structures produced by different parsers can be naturally compiled into forest, offering complementary strength to our single-view parser. Experimental results on benchmark data show that our method significantly outperforms the baseline supervised parser and other entire-tree based semi-supervised methods, such as self-training, co-training and tri-training.

Session 3E: Language Resources and Evaluation I
Harborside D/E   *Chair: Fei Xia*

A Robust Approach to Aligning Heterogeneous Lexical Resources
Mohammad Taher Pilehvar and Roberto Navigli   3:30–3:55

Lexical resource alignment has been an active field of research over the last decade. However, prior methods for aligning lexical resources have been either specific to a particular pair of resources, or heavily dependent on the availability of hand-crafted alignment data for the pair of resources to be aligned. Here we present a unified approach that can be applied to an arbitrary pair of lexical resources, including machine-readable dictionaries with no network structure. Our approach leverages a similarity measure that enables the structural comparison of senses across lexical resources, achieving state-of-the-art performance on the task of aligning WordNet to three different collaborative resources: Wikipedia, Wiktionary and OmegaWiki.

Predicting the relevance of distributional semantic similarity with contextual information
Philippe Muller, Cécile Fabre, and Clémentine Adam   3:55–4:20

Using distributional analysis methods to compute semantic proximity links between words has become commonplace in NLP. The resulting relations are often noisy or difficult to interpret in general. This paper focuses on the issues of evaluating a distributional resource and filtering the relations it contains, but instead of considering it in abstracto, we focus on pairs of words in context. In a discourse, we are interested in knowing if the semantic link between two items is a by-product of textual coherence or is irrelevant. We first set up a human annotation of semantic links with or without contextual information to show the importance of the textual context in evaluating the relevance of semantic similarity, and to assess the prevalence of actual semantic relations between word tokens. We then built an experiment to automatically predict this relevance, evaluated on the reliable reference data set which was the outcome of the first annotation. We show that in-document information greatly improve the prediction made by the similarity level alone.

[TACL] Temporal Annotation in the Clinical Domain
William Styler, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, and James Pustejovsky   4:20–4:45

This article discusses the requirements of a formal specification for the annotation of tem- poral information in clinical narratives. We discuss the implementation and extension of ISO-TimeML for annotating a corpus of clini- cal notes, known as the XXX corpus. To reflect the information task and the heavily inference-based reasoning demands in the domain, a new annotation guideline has been developed, "the XXX Guidelines to ISO-TimeML (XXX- TimeML)". To clarify what relations merit an- notation, we distinguish between linguistically- derived and inferentially-derived temporal or- derings in the text. We also apply a top per- forming TempEval 2013 system against this new resource to measure the difficulty of adapt- ing systems to the clinical domain. The corpus is available to the community and has been proposed for use in a SemEval 2015 task.

Learning Ensembles of Structured Prediction Rules
Corinna Cortes, Vitaly Kuznetsov, and Mehryar Mohri   5:00–6:00

We present a series of algorithms with theoretical guarantees for learning accurate ensembles of several structured prediction rules for which no prior knowledge is assumed. This includes a number of randomized and deterministic algorithms devised by converting on-line learning algorithms to batch ones, and a boosting-style algorithm applicable in the context of structured prediction with a large number of labels. We also report the results of extensive experiments with these algorithms. This is joint work with Vitaly Kuznetsov, NYU, and Mehryar Mohri, NYU/Google Research.

Keynote Address: Corinna Cortes

"Learning Ensembles of Structured Prediction Rules"

Monday, June 23, 2014, 5:00–6:00pm

Grand Ballroom V–VI

Abstract: We present a series of algorithms with theoretical guarantees for learning accurate ensembles of several structured prediction rules for which no prior knowledge is assumed. This includes a number of randomized and deterministic algorithms devised by converting on-line learning algorithms to batch ones, and a boosting-style algorithm applicable in the context of structured prediction with a large number of labels. We also report the results of extensive experiments with these algorithms.

This is joint work with Vitaly Kuznetsov, NYU, and Mehryar Mohri, NYU/Google Research.

Biography: Corinna Cortes is the Head of Google Research, NY, where she is working on a broad range of theoretical and applied large-scale machine learning problems. Prior to Google, Corinna spent more than ten years at AT&T Labs - Research, formerly AT&T Bell Labs, where she held a distinguished research position. Corinna's research work is well-known in particular for her contributions to the theoretical foundations of support vector machines (SVMs), for which she jointly with Vladimir Vapnik received the 2008 Paris Kanellakis Theory and Practice Award, and her work on data-mining in very large data sets for which she was awarded the AT&T Science and Technology Medal in the year 2000. Corinna received her MS degree in Physics from University of Copenhagen and joined AT&T Bell Labs as a researcher in 1989. She received her Ph.D. in computer science from the University of Rochester in 1993. Corinna is also a competitive runner.

Student Research Workshop Oral Highlights

This session provides students with a two-minute oral presentation for their poster. It includes all papers submitted to the student research workshop, as well as TACL, long, and short papers accepted to the general conference (and sponsored by the SRW).

A Mapping-Based Approach for General Formal Human Computer Interaction Using Natural Language

Vincent Letard, Sophie Rosset, and Gabriel Illouz   6:05–6:10

We consider the problem of mapping natural language written utterances expressing operational instructions to formal language expressions, applied to French and the R programming language. Developing a learning operational assistant requires the means to train and evaluate it, that is, a baseline system able to interact with the user. After presenting the guidelines of our work, we propose a model to represent the problem and discuss the fit of direct mapping methods to our task. Finally, we show that, while not resulting in excellent scores, a simple approach seems to be sufficient to provide a baseline for an interactive learning system.

An Exploration of Embeddings for Generalized Phrases

Wenpeng Yin and Hinrich Schütze   6:10–6:15

Deep learning embeddings have been successfully used for many natural language processing problems. Embeddings are mostly computed for word forms although lots of recent papers have extended this to other linguistic units like morphemes and word sequences. In this paper, we define the concept of generalized phrase that includes conventional linguistic phrases as well as skip-bigrams. We compute embeddings for generalized phrases and show in experimental evaluations on coreference resolution and paraphrase identification that such embeddings perform better than word form embeddings.

Learning Grammar with Explicit Annotations for Subordinating Conjunctions

Dongchen Li, Xiantao Zhang, and Xihong Wu   6:15–6:20

Data-driven approach for parsing may suffer from data sparsity when entirely unsupervised. External knowledge has been shown to effectively help mitigate this problem. Subordinating conjunctions impose important constraints on Chinese syntactic structures. This paper proposes a method to train a grammar with hierarchical subordinating conjunctions category knowledge as explicit annotations. Firstly, each part-of-speech tag of the subordinating conjunctions is annotated with the coarsest category in the hierarchical knowledge. These categories are human-defined to represent distinct syntactic constraints, and provide a good starting point for splitting. Secondly, based on the data-driven state-split approach, we establish a mapping from each automatic refined subcategory to one subcategory in the hierarchical knowledge. Then the data-driven splitting of these categories are restricted by the knowledge to avoid over refinement. Experiments show that constraining the grammar learning by the hierarchical knowledge improves parsing performance significantly on the baseline.

Going beyond sentences when applying tree kernels

Dmitry Ilvovsky   6:20–6:25

We go beyond the level of individual sentences applying parse tree kernels to paragraphs. We build a set of extended trees for a paragraph of text from the individual parse trees for sentences and learn short texts such as search results and social profile postings to take advantage of additional discourse-related information. Extension is based on coreferences and rhetoric structure relations between the phrases in different sentences. We evaluate our approach, tracking relevance classifi-cation improvement for multi-sentence search task. The search problem is formulated as classification of search results into the classes of relevant and irrelevant, learning from the Bing search results. We compare performances of individual sentence kernels with the ones for extended parse trees and show that adding discourse information to learning data helps to improve classification results.

Multi-document summarization using distortion-rate ratio

Ulukbek Attokurov and Ulug Bayazit   6:25–6:30

The current work adapts the optimal tree pruning algorithm introduced by Breiman et al. and extended by Chou et al. to the multi-document summarization task. The tree designed by Hierarchical Agglomerative Clustering algorithm is successively pruned with the optimal tree pruning algorithm and the sentences assigned to the leaves of the resultant pruned tree are included in the summary.

Disambiguating prepositional phrase attachment sites with sense information captured in contextualized distributional data

Clayton Greenberg 6:30–6:35

This work presents a supervised prepositional phrase (PP) attachment disambiguation system that uses contextualized distributional information as the distance metric for a nearest-neighbor classifier. Contextualized word vectors constructed from the GigaWord Corpus provide a method for implicit Word Sense Disambiguation (WSD), whose reliability helps this system outperform baselines and achieve comparable results to those of systems with full WSD modules. This suggests that targeted WSD methods are preferable to ignoring sense information and also to implementing WSD as an independent module in a pipeline.

Open Information Extraction for Spanish Language based on Syntactic Constraints

Alisa Zhila and Alexander Gelbukh 6:35–6:40

Open Information Extraction (Open IE) serves for the analysis of vast amounts of texts by extraction of assertions, or relations, in the form of tuples <argument 1; relation; argument 2>. Various approaches to Open IE have been designed to perform in a fast, unsupervised manner. All of them require language specific information for their implementation. In this work, we introduce an approach to Open IE based on syntactic constraints over POS tag sequences targeted at Spanish language. We describe the rules specific for Spanish language constructions and their implementation in ExtrHech, an Open IE system for Spanish. We also discuss language-specific issues of implementation. We compare ExtrHech's performance with that of ReVerb, a similar Open IE system for English, on a parallel dataset and show that these systems perform at a very similar level. We also compare ExtrHech's performance on a dataset of grammatically correct sentences against its performance on a dataset of random texts extracted from the Web, drastically different in their quality from the first dataset. The latter experiment shows robustness of ExtrHech on texts from the Web.

Improving Text Normalization via Unsupervised Model and Discriminative Reranking

Chen Li and Yang Liu 6:40–6:45

Various models have been developed for normalizing informal text. In this paper, we propose two methods to improve normalization performance. First is an unsupervised approach that automatically identifies pairs of a non-standard token and proper word from a large unlabeled corpus. We use semantic similarity based on continuous word vector representation, together with other surface similarity measurement. Second we propose a reranking strategy to combine the results from different systems. This allows us to incorporate information that is hard to model in individual systems as well as consider multiple systems to generate a final rank for a test case. Both word- and sentence-level optimization schemes are explored in this study. We evaluate our approach on data sets used in prior studies, and demonstrate that our proposed methods perform better than the state-of-the-art systems.

Semi-Automatic Development of KurdNet, The Kurdish WordNet

Purya Aliabadi 6:40–6:45

Recently, we reported on our efforts to build the first prototype of KurdNet. In this proposal, we highlight the shortcomings of the current prototype and put forward a detailed plan to transform this prototype to a full-fledged lexical database for the Kurdish language.

The following papers are also included in the SRW oral highlights session but are primarily normal full or short paper presentations part of the main conference proceedings. Their abstracts can therefore be found in their respective sessions (see the page number following each entry):

Anchors Regularized: Adding Robustness and Extensibility to Scalable Topic-Modeling Algorithms

Thang Nguyen, Yuening Hu, and Jordan Boyd-Graber (p.35)

Aspect Extraction with Automated Prior Knowledge Learning

Zhiyuan Chen, Arjun Mukherjee, and Bing Liu (p.35)

Discovering Latent Structure in Task-Oriented Dialogues

Ke Zhai and Jason D Williams (p.23)

Representation Learning for Text-level Discourse Parsing
Yangfeng Ji and Jacob Eisenstein (p.23)

ConnotationWordNet: Learning Connotation over the Word+Sense Network
Jun Seok Kang, Song Feng, Leman Akoglu, and Yejin Choi (p.105)

ReNew: A Semi-Supervised Framework for Generating Domain-Specific Lexicons and Sentiment Analysis
Zhe Zhang and Munindar P. Singh (p.46)

A Study of Concept-based Weighting Regularization for Medical Records Search
Yue Wang, Xitong Liu, and Hui Fang (p.47)

Political Ideology Detection Using Recursive Neural Networks
Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik (p.81)

Simple Negation Scope Resolution through Deep Parsing: A Semantic Solution to a Semantic Problem
Woodley Packard, Emily M. Bender, Jonathon Read, Stephan Oepen, and Rebecca Dridan (p.24)

Fast Easy Unsupervised Domain Adaptation with Marginalized Structured Dropout
Yi Yang and Jacob Eisenstein (p.89)

[TACL] Heterogeneous Networks and Their Applications: Scientometrics, Name Disambiguation, and Topic Modeling
Ben King, Rahul Jha, and Dragomir R. Radev (p.44)

Linguistic Considerations in Automatic Question Generation
Karen Mazidi and Rodney D. Nielsen (p.78)

Learning Semantic Hierarchies via Word Embeddings
Ruiji Fu, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu (p.82)

Toward Better Chinese Word Segmentation for SMT via Bilingual Constraints
Xiaodong Zeng, Lidia S. Chao, Derek F. Wong, Isabel Trancoso, and Liang Tian (p.85)

Two Knives Cut Better Than One: Chinese Word Segmentation with Dual Decomposition
Mengqiu Wang, Rob Voigt, and Christopher Manning (p.55)

Posters: TACL papers

[TACL] Entity Linking meets Word Sense Disambiguation: a Unified Approach
Andrea Moro, Alessandro Raganato, and Roberto Navigli

Entity Linking (EL) and Word Sense Disambiguation (WSD) both address the lexical ambiguity of language. But while the two tasks are pretty similar, they differ in a fundamental aspect: in EL the textual mention can be linked to an entity which may or may not contain the exact mention, while in WSD there is a perfect match between the word form (better, its lemma) and a suitable sense. In this paper we present a unified graph-based approach to EL and WSD based on a loose identification of candidate meanings coupled with a densest subgraph heuristic which selects high-coherence semantic interpretations. Our experiments show state-of-the-art performances on both tasks on four different datasets, including a multilingual setting.

[TACL] Data-Driven Metaphor Recognition and Explanation
Hongsong Li, Kenny Q. Zhu, and Haixun Wang

Recognizing metaphors and identifying thesource-target mappings is an important taskas metaphorical text poses a big challenge formachine reading. To address this problem, weautomatically acquire a metaphor knowledgebase and an isA knowledge base from billionsof web pages. Using the knowledge bases, wedevelop an inferencing mechanism to recognizeand explain the metaphors in text. Toour knowledge, this is the first purely datadrivenapproach of probabilistic metaphor acquisition,recognition and explanation. Our resultsshow that it is effective and significantlyoutperforms other state-of-the-art methods.

[TACL] Grounded Compositional Semantics for Finding and Describing Images with Sentences
Richard Socher, Andrej Karpathy, Quoc V. Le, Christopher Manning, and Andrew Ng

Previous work on Recursive Neural Networks (RNNs) shows that these models can produce compositional feature vectors for accurately representing and classifying sentences or images. However, the sentence vectors of previous models cannot accurately represent visually grounded meaning. We introduce the DT-RNN model which uses dependency trees to embed sentences into a vector space in order to retrieve images that are described by those sentences. Unlike previous RNN-based models which use constituency trees, DT-RNNs naturally focus on the action and agents in a sentence. They are better able to abstract from the details of word order and syntactic expression. DT-RNNs outperform other recursive and recurrent neural networks, kernelized CCA and a bag-of-words baseline on the tasks of finding an image that fits a sentence description and vice versa. They also give more similar representations to sentences that describe the same image.

[TACL] Parallel Algorithms for Unsupervised Tagging
Sujith Ravi, Sergei Vassilivitskii, and Vibhor Rastogi

We propose a new method for unsupervised tagging that finds minimal models, that are further improved by Expectation Minimization training. In contrast to previous approaches that rely on manually specified and multi-step heuristics for model minimization, our approach is a simple greedy approximation algorithm DMLC (Distributed Minimum Label Cover) that solves this objective in a single step. We extend the method and show how to efficiently parallelize the algorithm on modern parallel computing platforms while preserving approximation guarantees. The new method easily scales to large data and grammar sizes, overcoming the memory bottleneck in previous approaches. We demonstrate the power of the new algorithm by evaluating on various sequence labeling tasks—Part-of-Speech tagging for multiple languages (including low-resource languages), with complete and incomplete dictionaries, and supertagging, a complex sequence labeling task, where the grammar size alone can grow to millions of entries. Our results show that for all of these settings, our method achieves state-of-the-art scalable performance that yields high quality tagging outputs.

[TACL] Heterogeneous Networks and Their Applications: Scientometrics, Name Disambiguation, and Topic Modeling
Ben King, Rahul Jha, and Dragomir R. Radev

In this paper, we present heterogeneous networks as a way to unify lexical networks with relational data. We build a unified ACL Anthology network, tying together the citation, author collaboration, and term-cooccurence networks with affiliation and venue relations. This representation proves to be convenient and allows problems such as name disambiguation, topic modeling, and the measurement of scientific impact to be easily solved using only this network and off-the-shelf graph algorithms.

[TACL] Discriminative Lexical Semantic Segmentation with Gaps: Running the MWE Gamut
Nathan Schneider, Emily Danchik, Chris Dyer, and Noah A. Smith

We present a novel representation, evaluation measure, and supervised models for the task of identifying the multiword expressions (MWEs) in a sentence, resulting in a lexical semantic segmentation. Our approach generalizes a standard chunking representation to encode a subset of projective MWEs containing gaps, thereby enabling efficient sequence tagging algorithms for feature-rich discriminative models. Experiments on a new dataset of English web text offer the first linguistically-driven evaluation of MWE identification with truly heterogeneous expression types. Our statistical sequence model greatly outperforms a lookup-based segmentation procedure, achieving 60% F1 for MWE identification.

Posters: Long papers

Interpretable Semantic Vectors from a Joint Model of Brain- and Text- Based Meaning
Alona Fyshe, Partha P. Talukdar, Brian Murphy, and Tom M. Mitchell

Vector space models (VSMs) represent word meanings as points in a high dimensional space. VSMs are typically created using a large text corpora, and so represent word semantics as observed in text. We present a new algorithm (JNNSE) that can incorporate a measure of semantics not previously used to create VSMs: brain activation data recorded while people read words. The resulting model takes advantage of the complementary strengths and weaknesses of corpus and brain activation data to give a more complete representation of semantics. Evaluations show that the model 1) matches a behavioral measure of semantics more closely, 2) can be used to predict corpus data for unseen words and 3) has predictive power that generalizes across brain imaging technologies and across subjects. We believe that the model is thus a more faithful representation of mental vocabularies.

Single-Agent vs. Multi-Agent Techniques for Concurrent Reinforcement Learning of Negotiation Dialogue Policies
Kallirroi Georgila, Claire Nelson, and David Traum

We use single-agent and multi-agent Reinforcement Learning (RL) for learning dialogue policies in a resource allocation negotiation scenario. Two agents learn concurrently by interacting with each other without any need for simulated users (SUs) to train against or corpora to learn from. In particular, we compare the Q-learning, Policy Hill-Climbing (PHC) and Win or Learn Fast Policy Hill-Climbing (PHC-WoLF) algorithms, varying the scenario complexity (state space size), the number of training episodes, the learning rate, and the exploration rate. Our results show that generally Q-learning fails to converge whereas PHC and PHC-WoLF always converge and perform similarly. We also show that very high gradually decreasing exploration rates are required for convergence. We conclude that multi-agent RL of dialogue policies is a promising alternative to using single-agent RL and SUs or learning directly from corpora.

A Linear-Time Bottom-Up Discourse Parser with Constraints and Post-Editing
Vanessa Wei Feng and Graeme Hirst

Text-level discourse parsing remains a challenge. The current state-of-the-art overall accuracy in relation assignment is 55.73%, achieved by Joty et al.(2013). However, their model has a high order of time complexity, and thus cannot be applied in practice. In this work, we develop a much faster model whose time complexity is linear in the number of sentences. Our model adopts a greedy bottom-up approach, with two linear-chain CRFs applied in cascade as local classifiers. To enhance the accuracy of the pipeline, we add additional constraints in the Viterbi decoding of the first CRF. In addition to efficiency, our parser also significantly outperforms the state of the art. Moreover, our novel approach of post-editing, which modifies a fully-built tree by considering information from constituents on upper levels, can further improve the accuracy.

Negation Focus Identification with Contextual Discourse Information
Bowei Zou, Guodong Zhou, and Qiaoming Zhu

Negative expressions are common in natural language text and play a critical role in information extraction. However, the performances of current systems are far from satisfaction, largely due to its focus on intra-sentence information and its failure to consider inter-sentence information. In this paper, we propose a graph model to enrich intra-sentence features with inter-sentence features from both lexical and topic perspectives. Evaluation on the *SEM 2012 shared task corpus indicates the usefulness of contextual discourse information

in negation focus identification and justifies the effectiveness of our graph model in capturing such global information.

New Word Detection for Sentiment Analysis
Minlie Huang, Borui Ye, Yichen Wang, Haiqiang Chen, Junjun Cheng, and Xiaoyan Zhu

Automatic extraction of new words is an indispensable precursor to many NLP tasks such as Chinese word segmentation, named entity extraction, and sentiment analysis. This paper aims at extracting new sentiment words from large-scale user-generated content. We propose a fully unsupervised, purely data-driven framework for this purpose. We design statistical measures respectively to quantify the utility of a lexical pattern and to measure the possibility of a word being a new word. The method is almost free of linguistic resources (except POS tags), and requires no elaborated linguistic rules. We also demonstrate how new sentiment word will benefit sentiment analysis. Experiment results demonstrate the effectiveness of the proposed method.

ReNew: A Semi-Supervised Framework for Generating Domain-Specific Lexicons and Sentiment Analysis
Zhe Zhang and Munindar P. Singh

The sentiment captured in opinionated text provides interesting and valuable information for social media services. However, due to the complexity and diversity of linguistic representations, it is challenging to build a framework that accurately extracts such sentiment. We propose a semi-supervised framework for generating a domain-specific sentiment lexicon and inferring sentiments at the segment level. Our framework can greatly reduce the human effort for building a domain-specific sentiment lexicon with high quality. Specifically, in our evaluation, working with just 20 manually labeled reviews, it generates a domain-specific sentiment lexicon that yields weighted average F-Measure gains of 3%. Our sentiment classification model achieves approximately 1% greater accuracy than a state-of-the-art approach based on elementary discourse units.

A Decision-Theoretic Approach to Natural Language Generation
Nathan McKinley and Soumya Ray

We study the problem of generating an English sentence given an underlying probabilistic grammar, a world and a communicative goal. We model the generation problem as a Markov decision process with a suitably defined reward function that reflects the communicative goal. We then use probabilistic planning to solve the MDP and generate a sentence that, with high probability, accomplishes the communicative goal. We show empirically that our approach can generate complex sentences with a speed that generally matches or surpasses the state of the art. Further, we show that our approach is anytime and can handle complex communicative goals, including negated goals.

Generating Code-switched Text for Lexical Learning
Igor Labutov and Hod Lipson

A vast majority of L1 vocabulary acquisition occurs through incidental learning during reading. We propose a probabilistic approach to generating code-mixed text as an L2 technique for increasing retention in adult lexical learning through reading. Our model that takes as input a bilingual dictionary and an English text, and generates a code-switched text that optimizes a defined "learnability" metric by constructing a factor graph over lexical mentions. Using an artificial language vocabulary, we evaluate a set of algorithms for generating code-switched text automatically by presenting it to Mechanical Turk subjects and measuring recall in a sentence completion task.

Omni-word Feature and Soft Constraint for Chinese Relation Extraction
Yanping Chen, Qinghua Zheng, and Wei Zhang

Chinese language is an ancient hieroglyphic. It is inattentive to structure. Therefore, segmenting and parsing Chinese are more difficult and less accurate. In this paper, we propose an Omni-word feature and a soft constraint method for Chinese relation extraction. The Omni-word feature uses every potential word in a sentence as lexicon feature, reducing errors caused by word segmentation. In order to utilize the structure information of a relation instance, we discuss how soft constraint can be used to capture the local dependency. Both Omni-word feature and soft constraint make a better use of sentence information and minimize the influences caused by Chinese word segmentation and parsing. We test these methods on the ACE 2005 RDC Chinese corpus. The results show a significant improvement in Chinese relation extraction, outperforming other methods in F-score by 10% in 6 relation types and 15% in 18 relation subtypes.

Bilingual Active Learning for Relation Classification via Pseudo Parallel Corpora
Longhua Qian, Haotian Hui, Ya'nan Hu, Guodong Zhou, and Qiaoming Zhu

Active learning (AL) has been proven effective to reduce human annotation efforts in NLP. However, previous studies on AL are limited to applications in a single language. This paper proposes a bilingual active learning paradigm for relation classification, where the unlabeled instances are first jointly chosen in terms of their prediction uncertainty scores in two languages and then manually labeled by an oracle. Instead of using a parallel corpus, labeled and unlabeled instances in one language are translated into ones in the other language and all instances in both languages are then fed into a bilingual active learning engine as pseudo parallel corpora. Experimental results on the ACE RDC 2005 Chinese and English corpora show that bilingual active learning for relation classification significantly outperforms monolingual active learning.

Learning Soft Linear Constraints with Application to Citation Field Extraction
Sam Anzaroot, Alexandre Passos, David Belanger, and Andrew McCallum

Accurately segmenting a citation string into fields for authors, titles, etc. is a challenging task because the output typically obeys various global constraints. Previous work has shown that modeling soft constraints, where the model is encouraged, but not require to obey the constraints, can substantially improve segmentation performance. On the other hand, for imposing hard constraints, dual decomposition is a popular technique for efficient prediction given existing algorithms for unconstrained inference. We extend dual decomposition to perform prediction subject to soft constraints. Moreover, with a technique for performing inference given soft constraints, it is easy to automatically generate large families of constraints and learn their costs with a simple convex optimization problem during training. This allows us to obtain substantial gains in accuracy on a new, challenging citation extraction dataset.

A Study of Concept-based Weighting Regularization for Medical Records Search
Yue Wang, Xitong Liu, and Hui Fang

An important search task in the biomedical domain is to find medical records of patients who are qualified for a clinical trial. One commonly used approach is to apply NLP tools to map terms from queries and documents to concepts and then compute the relevance scores based on the concept-based representation. However,the mapping results are not perfect, and none of previous work studied how to deal with them in the retrieval process. In this paper, we focus on addressing the limitations caused by the imperfect mapping results and study how to further improve the retrieval performance of the concept-based ranking methods. In particular, we apply axiomatic approaches and propose two weighting regularization methods that adjust the weighting based on the relations among the concepts. Experimental results show that the proposed methods are effective to improve the retrieval performance, and their performances are comparable to other top-performing systems in the TREC Medical Records Track.

Learning to Predict Distributions of Words Across Domains
Danushka Bollegala, David Weir, and John Carroll

Although the distributional hypothesis has been applied successfully in many natural language processing tasks, systems using distributional information have been limited to a single domain because the distribution of a word can vary between domains as the word's predominant meaning changes. However, if it were possible to predict how the distribution of a word changes from one domain to another, the predictions could be used to adapt a system trained in one domain to work in another. We propose an unsupervised method to predict the distribution of a word in one domain, given its distribution in another domain. We evaluate our method on two tasks: cross-domain part-of-speech tagging and cross-domain sentiment classification. In both tasks, our method significantly outperforms competitive baselines and returns results that are statistically comparable to current state-of-the-art methods, while requiring no task-specific customisations.

How to make words with vectors: Phrase generation in distributional semantics
Georgiana Dinu and Marco Baroni

We introduce the problem of generation in distributional semantics: Given a distributional vector representing some meaning, how can we generate the phrase that best expresses that meaning? We motivate this novel challenge on theoretical and practical grounds and propose a simple data-driven approach to the estimation of generation functions. We test this in a monolingual scenario (paraphrase generation) as well as in a cross-lingual setting (translation by synthesizing adjective-noun phrase vectors in English and generating the equivalent expressions in Italian).

Vector space semantics with frequency-driven motifs
Shashank Srivastava and Eduard Hovy

Traditional models of distributional semantics suffer from computational issues such as data sparsity for individual lexemes and complexities of modeling semantic composition when dealing with structures larger than single lexical items. In this work, we present a frequency-driven paradigm for robust distributional semantics in terms of semantically cohesive lineal constituents, or motifs. The framework subsumes issues such as differential compositional as well as non-compositional behavior of phrasal consituents, and circumvents some problems of data sparsity by design. We design a segmentation model to optimally partition a sentence into lineal constituents, which can be used to define distributional contexts that are less noisy, semantically more interpretable, and linguistically disambiguated. Hellinger PCA embeddings learnt using the framework show competitive results on empirical tasks.

Lexical Inference over Multi-Word Predicates: A Distributional Approach
Omri Abend, Shay B. Cohen, and Mark Steedman

Representing predicates in terms of their argument distribution is common practice in NLP. Multi-word predicates (MWPs) in this context are often either disregarded or considered as fixed expressions. The latter treatment is unsatisfactory in two ways: (1) identifying MWPs is notoriously difficult, (2) MWPs show varying degrees of compositionality and could benefit from taking into account the identity of their component parts. We propose a novel approach that integrates the distributional representation of multiple sub-sets of the MWP's words. We assume a latent distribution over sub-sets of the MWP, and estimate it relative to a downstream prediction task. Focusing on the supervised identification of lexical inference relations, we compare against state-of-the-art baselines that consider a single sub-set of an MWP, obtaining substantial improvements. To our knowledge, this is the first work to address lexical relations between MWPs of varying degrees of compositionality within distributional semantics.

A Convolutional Neural Network for Modelling Sentences
Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom

The ability to accurately represent sentences is central to language understanding. We describe a convolutional architecture dubbed the Dynamic Convolutional Neural Network (DCNN) that we adopt for the semantic modelling of sentences. The network uses Dynamic k-Max Pooling, a global pooling operation over linear sequences. The network handles input sentences of varying length and induces a feature graph over the sentence that is capable of explicitly capturing short and long-range relations. The network does not rely on a parse tree and is easily applicable to any language. We test the DCNN in four experiments: small scale binary and multi-class sentiment prediction, six-way question classification and Twitter sentiment prediction by distant supervision. The network achieves excellent performance in the first three tasks and a greater than 25% error reduction in the last task with respect to the strongest baseline.

Online Learning in Tensor Space
Yuan Cao and Sanjeev Khudanpur

We propose an online learning algorithm based on tensor-space models. A tensor-space model represents data in a compact way, and via rank-1 approximation the weight tensor can be made highly structured hence the number of parameters to be estimated is significantly smaller than the traditional vector-space models. This regularizes the model complexity and makes the tensor model highly effective under situations where a large feature set is defined but very low resource is available for training. We experimented the proposed algorithm on a parsing task, and our results show that even with very little training data the learning algorithm based on tensor model performs well, and gives significantly better results than learning algorithms based on traditional vector-space models.

Graph-based Semi-Supervised Learning of Translation Models from Monolingual Data
Avneesh Saluja, Hany Hassan, Kristina Toutanova, and Chris Quirk

Statistical phrase-based translation learns translation rules from bilingual corpora, and has traditionally only used monolingual evidence to construct features that rescore existing translation candidates. In this work, we present a semi-supervised graph-based approach for generating new translation rules that leverages bilingual and monolingual data. The proposed technique first constructs phrase graphs using both source and target language monolingual corpora. Next, graph propagation identifies translations of phrases that were not observed in the bilingual corpus, assuming that similar phrases have similar translations. We report results on a large Arabic-English system and a medium-sized Urdu-English system. Our proposed approach significantly

improves the performance of competitive phrase-based systems, leading to consistent improvements between 1 and 4 BLEU points on standard evaluation sets.

Using Discourse Structure Improves Machine Translation Evaluation
Francisco Guzmán, Shafiq Joty, Lluís Màrquez, and Preslav Nakov

We present experiments in using discourse structure for improving machine translation evaluation. We first design two discourse-aware similarity measures, which use all-subtree kernels to compare discourse parse trees in accordance with the Rhetorical Structure Theory. Then, we show that these measures can help improve a number of existing machine translation evaluation metrics both at the segment- and at the system-level. Rather than proposing a single new metric, we show that discourse information is complementary to the state-of-the-art evaluation metrics, and thus should be taken into account in the development of future richer evaluation metrics.

Learning Continuous Phrase Representations for Translation Modeling
Jianfeng Gao, Xiaodong He, Wen-tau Yih, and Li Deng

This paper tackles the sparsity problem in estimating phrase translation probabilities by learning continuous phrase representations, whose distributed nature enables the sharing of related phrases in their representations. A pair of source and target phrases are projected into continuous-valued vector representations in a low-dimensional latent space, where their translation score is computed by the distance between the pair in this new space. The projection is performed by a neural network whose weights are learned on parallel training data. Experimental evaluation has been performed on two WMT translation tasks. Our best result improves the performance of a state-of-the-art phrase-based statistical machine translation system trained on WMT 2012 French-English data by up to 1.3 BLEU points.

Adaptive Quality Estimation for Machine Translation
Marco Turchi, Antonios Anastasopoulos, José G. C. de Souza, and Matteo Negri

The automatic estimation of machine translation (MT) output quality is a hard task in which the selection of the appropriate algorithm and the most predictive features over reasonably sized training sets plays a crucial role. When moving from controlled lab evaluations to real-life scenarios the task becomes even harder. For current MT quality estimation (QE) systems, additional complexity comes from the difficulty to model user and domain changes. The instability of the systems with respect to data coming from different distributions, in fact, calls for adaptive solutions that react to new operating conditions. To tackle this issue we propose an online framework for adaptive QE that targets reactivity and robustness to user and domain changes. Contrastive experiments in different testing conditions involving user and domain changes demonstrate the effectiveness of our approach.

Learning Grounded Meaning Representations with Autoencoders
Carina Silberer and Mirella Lapata

In this paper we address the problem of grounding distributional representations of lexical meaning. We introduce a new model which uses stacked autoencoders to learn higher-level embeddings from textual and visual input. The two modalities are encoded as vectors of attributes and are obtained automatically from text and images, respectively. We evaluate our model on its ability to simulate similarity judgments and concept categorization. On both tasks, our approach outperforms baselines and related models.

Joint POS Tagging and Transition-based Constituent Parsing in Chinese with Non-local Features
Zhiguo Wang and Nianwen Xue

We propose three improvements to address the drawbacks of state-of-the-art transition-based constituent parsers. First, to resolve the error propagation problem of the traditional pipeline approach, we incorporate POS tagging into the syntactic parsing process. Second, to alleviate the negative influence of size differences among competing action sequences, we align parser states during beam-search decoding. Third, to enhance the power of parsing models, we enlarge the feature set with non-local features and semi-supervised word cluster features. Experimental results show that these modifications improve parsing performance significantly. Evaluated on the Chinese TreeBank (CTB), our final performance reaches 86.3% (F1) when trained on CTB 5.1, and 87.1% when trained on CTB 6.0, and these results outperform all state-of-the-art parsers.

Strategies for Contiguous Multiword Expression Analysis and Dependency Parsing
Marie Candito and Matthieu Constant

In this paper, we investigate various strategies to predict both syntactic dependency parsing and contiguous multiword expression (MWE) recognition, testing them on the dependency version of French Treebank, as instantiated in the SPMRL Shared Task. Our work focuses on using an alternative representation of syntactically regular MWEs, which captures their syntactic internal structure. We obtain a system with comparable performance to that of previous works on this dataset, but which predicts both syntactic dependencies and the internal structure of MWEs. This can be useful for capturing the various degrees of semantic compositionality of MWEs.

Correcting Preposition Errors in Learner English Using Error Case Frames and Feedback Messages

Ryo Nagata, Mikko Vilenius, and Edward Whittaker

This paper presents a novel framework called error case frames for correcting preposition errors. They are case frames specially designed for describing and correcting preposition errors. Their most distinct advantage is that they can correct errors with feedback messages explaining why the preposition is erroneous. This paper proposes a method for automatically generating them by comparing learner and native corpora. Experiments show that error correction with feedback messages provided by error case frames are effective in language learning assistance, achieving a correction performance comparable to conventional methods.

Posters: Short papers

Exploring the Relative Role of Bottom-up and Top-down Information in Phoneme Learning

Abdellah Fourtassi, Thomas Schatz, Balakrishnan Varadarajan, and Emmanuel Dupoux

We test both bottom-up and top-down approaches in learning the phonemic status of the sounds of English and Japanese. We used large corpora of spontaneous speech to provide the learner with an input that models both the linguistic properties and statistical regularities of each language. We found both approaches to help discriminate between allophonic and phonemic contrasts with a high degree of accuracy, although top-down cues proved to be effective only on an interesting subset of the data.

Biases in Predicting the Human Language Model

Alex B. Fine, Austin F. Frank, T. Florian Jaeger, and Benjamin Van Durme

We consider the prediction of three human behavioral measures — lexical decision, word naming, and picture naming — through the lens of domain bias in language modeling. Contrasting the predictive ability of statistics derived from 6 different corpora, we find intuitive results showing that, e.g., a British corpus over-predicts the speed with which an American will react to the words ward and duke, and that the Google n-grams over-predicts familiarity with technology terms. This study aims to provoke increased consideration of the human language model by NLP practitioners: biases are not limited to differences between corpora (i.e. "train" vs. "test"); they can exist as well between corpora and the intended user of the resultant technology.

Probabilistic Labeling for Efficient Referential Grounding based on Collaborative Discourse

Changsong Liu, Lanbo She, Rui Fang, and Joyce Y. Chai

When humans and artificial agents (e.g. robots) have mismatched perceptions of the shared environment, referential communication between them becomes difficult. To mediate perceptual differences, this paper presents a new approach using probabilistic labeling for referential grounding. This approach aims to integrate different types of evidence from the collaborative referential discourse into a unified scheme. Its probabilistic labeling procedure can generate multiple grounding hypotheses to facilitate follow-up dialogue. Our empirical results have shown the probabilistic labeling approach significantly outperforms a previous graph-matching approach for referential grounding.

A Composite Kernel Approach for Dialog Topic Tracking with Structured Domain Knowledge from Wikipedia

Seokhwan Kim, Rafael E. Banchs, and Haizhou Li

Dialog topic tracking aims at analyzing and maintaining topic transitions in ongoing dialogs. This paper proposes a composite kernel approach for dialog topic tracking to utilize various types of domain knowledge obtained from Wikipedia. Two kernels are defined based on history sequences and context trees constructed based on the extracted features. The experimental results show that our composite kernel approach can signif-

icantly improve the performances of topic tracking in mixed-initiative human-human dialogs.

An Extension of BLANC to System Mentions
Xiaoqiang Luo, Sameer Pradhan, Marta Recasens, and Eduard Hovy

BLANC is a link-based coreference evaluation metric for measuring the quality of coreference systems on gold mentions. This paper extends the original BLANC ("BLANC-gold" henceforth) to system mentions, removing the gold mention assumption. The proposed BLANC falls back seamlessly to the original one if system mentions are identical to gold mentions, and it is shown to strongly correlate with existing metrics on the 2011 and 2012 CoNLL data.

Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation
Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng, and Michael Strube

The definitions of two coreference scoring metrics—B3 and CEAF—are underspecified with respect to predicted, as opposed to key (or gold) mentions. Several variations have been proposed that manipulate either, or both, the key and predicted mentions in order to get a one-to-one mapping. On the other hand, the metric BLANC was, until recently, limited to scoring partitions of key mentions. In this paper, we (i) argue that mention manipulation for scoring predicted mentions is unnecessary, and potentially harmful as it could produce unintuitive results; (ii) illustrate the application of all these measures to scoring predicted mentions; (iii) make available an open-source, thoroughly-tested reference implementation of the main coreference evaluation measures; and (iv) rescore the results of the CoNLL-2011/2012 shared task systems with this implementation. This will help the community accurately measure and compare new end-to-end coreference resolution algorithms.

Measuring Sentiment Annotation Complexity of Text
Aditya Joshi, Abhijit Mishra, Nivvedan Senthamilselvan, and Pushpak Bhattacharyya

The effort required for a human annotator to detect sentiment is not uniform for all texts, irrespective of his/her expertise. We aim to predict a score that quantifies this effort, using linguistic properties of the text. Our proposed metric is called Sentiment Annotation Complexity (SAC). As for training data, since any direct judgment of complexity by a human annotator is fraught with subjectivity, we rely on cognitive evidence from eye-tracking. The sentences in our dataset are labeled with SAC scores derived from eye-fixation duration. Using linguistic features and annotated SACs, we train a regressor that predicts the SAC with a best mean error rate of 22.02% for five-fold cross-validation. We also study the correlation between a human annotator's perception of complexity and a machine's confidence in polarity determination. The merit of our work lies in (a) deciding the sentiment annotation cost in, for example, a crowdsourcing setting,(b) choosing the right classifier for sentiment prediction.

Improving Citation Polarity Classification with Product Reviews
Charles Jochim and Hinrich Schütze

Recent work classifying citations in scientific literature has shown that it is possible to improve classification results with extensive feature engineering. While this result confirms that citation classification is feasible, there are two drawbacks to this approach: (i) it requires a large annotated corpus for supervised classification, which in the case of scientific literature is quite expensive; and (ii) feature engineering that is too specific to one area of scientific literature may not be portable to other domains, even within scientific literature. In this paper we address these two drawbacks. First, we frame citation classification as a domain adaptation task and leverage the abundant labeled data available in other domains. Then, to avoid over-engineering specific citation features for a particular scientific domain, we explore a deep learning neural network approach that has shown to generalize well across domains using unigram and bigram features. We achieve better citation classification results with this cross-domain approach than using in-domain classification.

Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification
Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu

We propose Adaptive Recursive Neural Network (AdaRNN) for target-dependent Twitter sentiment classification. AdaRNN adaptively propagates the sentiments of words to target depending on the context and syntactic relationships between them. It consists of more than one composition functions, and we model the adaptive sentiment propagations as distributions over these composition functions. The experimental studies illustrate that AdaRNN improves the baseline methods. Furthermore, we introduce a manually annotated dataset for target-dependent Twitter sentiment analysis.

Sprinkling Topics for Weakly Supervised Text Classification
Swapnil Hingmire and Sutanu Chakraborti

Supervised text classification algorithms require a large number of documents labeled by humans, that involve a labor-intensive and time consuming process. In this paper, we propose a weakly supervised algorithm in which supervision comes in the form of labeling of Latent Dirichlet Allocation (LDA) topics. We then use this weak supervision to "sprinkle" artificial words to the training documents to identify topics in accordance with the underlying class structure of the corpus based on the higher order word associations. We evaluate this approach to improve performance of text classification on three real world datasets.

A Feature-Enriched Tree Kernel for Relation Extraction
Le Sun and Xianpei Han

Tree kernel is an effective technique for relation extraction. However, the traditional syntactic tree representation is often too coarse or ambiguous to accurately capture the semantic relation information between two entities. In this paper, we propose a new tree kernel, called feature-enriched tree kernel (FTK), which can enhance the traditional tree kernel by: 1) refining the syntactic tree representation by an-notating each tree node with a set of discriminant features; and 2) proposing a new tree kernel which can better measure the syntactic tree similarity by taking all features into consideration. Experimental results show that our method can achieve a 5.4% F-measure improvement over the traditional convolution tree kernel.

Employing Word Representations and Regularization for Domain Adaptation of Relation Extraction
Thien Huu Nguyen and Ralph Grishman

Relation extraction suffers from a performance loss when a model is applied to out-of-domain data. This has fostered the development of domain adaptation techniques for relation extraction. This paper evaluates word embeddings and clustering on adapting feature-based relation extraction systems. We systematically explore various ways to apply word embeddings and show the best adaptation improvement by combining word cluster and word embedding information. Finally, we demonstrate the effectiveness of regularization for the adaptability of relation extractors.

Graph Ranking for Collective Named Entity Disambiguation
Ayman Alhelbawy and Robert Gaizauskas

Named Entity Disambiguation (NED)refers to the task of mapping different named entity mentions in running text to their correct interpretations in a specific knowledge base (KB). This paper presents a collective disambiguation approach using a graph model. All possible NE candidates are represented as nodes in the graph and associations between different candidates are represented by edges between the nodes. Each node has an initial confidence score, e.g. entity popularity. Page-Rank is used to rank nodes and the final rank is combined with the initial confidence for candidate selection. Experiments on 27819 NE textual mentions show the effectiveness of using Page-Rank in conjunction with initial confidence: 87% accuracy is achieved, outperforming both baseline and state-of-the-art approaches.

Descending-Path Convolution Kernel for Syntactic Structures
Chen Lin, Timothy Miller, Alvin Kho, Steven Bethard, Dmitriy Dligach, Sameer Pradhan, and Guergana Savova

Convolution tree kernels are an efficient and effective method for comparing syntactic structures in NLP methods. However, current kernel methods such as subset tree kernel and partial tree kernel understate the similarity of very similar tree structures. Although soft-matching approaches can improve the similarity scores, they are corpus-dependent and match relaxations may be task-specific. We propose an alternative approach called descending path kernel which gives intuitive similarity scores on comparable structures. This method is evaluated on two temporal relation extraction tasks and demonstrates its advantage over rich syntactic representations.

Entities' Sentiment Relevance
Zvi Ben-Ami, Ronen Feldman, and Binyamin Rosenfeld

Sentiment relevance detection problems occur when there is a sentiment expression in a text, and there is the question of whether or not the expression is related to a given entity or, more generally, to a given situation. The paper discusses variants of the problem, and shows that it is distinct from other somewhat similar problems occurring in the field of sentiment analysis and opinion mining. We experimentally demonstrate that using the in-formation about relevancy significantly affects the final sentiment evaluation of the entities. We then compare a set of different algorithms for solving the relevance detection problem. The most accurate results

are achieved by algorithms that use certain document-level information about the target entities. We show that this information can be accurately extracted using supervised classification methods.

Automatic Detection of Multilingual Dictionaries on the Web
Gintare Grigonyte and Timothy Baldwin

This paper presents an approach to query construction to detect multilingual dictionaries for predetermined language combinations on the web, based on the identification of terms which are likely to occur in bilingual dictionaries but not in general web documents. We use eight target languages for our case study, and train our method on pre-identified multilingual dictionaries and the Wikipedia dump for each of our languages.

Automatic Detection of Cognates Using Orthographic Alignment
Alina Maria Ciobanu and Liviu P. Dinu

Words undergo various changes when entering new languages. Based on the assumption that these linguistic changes follow certain rules, we propose a method for automatically detecting pairs of cognates employing an orthographic alignment method which proved relevant for sequence alignment in computational biology. We use aligned subsequences as features for machine learning algorithms in order to infer rules for linguistic changes undergone by words when entering new languages and to discriminate between cognates and non-cognates. Given a list of known cognates, our approach does not require any other linguistic information. However, it can be customized to integrate historical information regarding language evolution.

Automatically constructing Wordnet Synsets
Khang Nhut Lam, Feras Al Tarouti, and Jugal Kalita

Manually constructing a Wordnet is a difficult task, needing years of experts' time. As a first step to automatically construct full Wordnets, we propose approaches to generate Wordnet synsets for languages both resource-rich and resource-poor, using publicly available Wordnets, a machine translator and/or a single bilingual dictionary. Our algorithms translate synsets of existing Wordnets to a target language T, then apply a ranking method on the translation candidates to find best translations in T. Our approaches are applicable to any language which has at least one existing bilingual dictionary translating from English to it.

Constructing a Turkish-English Parallel TreeBank
Olcay Taner Yıldız, Ercan Solak, Onur Görgün, and Razieh Ehsani

In this paper, we report our preliminary efforts in building an English-Turkish parallel treebank corpus for statistical machine translation. In the corpus, we manually generated parallel trees for about 5,000 sentences from Penn Treebank. English sentences in our set have a maximum of 15 tokens, including punctuation. We constrained the translated trees to the reordering of the children and the replacement of the leaf nodes with appropriate glosses. We also report the tools that we built and used in our tree translation task.

Improved Typesetting Models for Historical OCR
Taylor Berg-Kirkpatrick and Dan Klein

We present richer typesetting models that extend the unsupervised historical document recognition system of Berg-Kirkpatrick et al. (2013). The first model breaks the independence assumption between vertical offsets of neighboring glyphs and, in experiments, substantially decreases transcription error rates. The second model simultaneously learns multiple font styles and, as a result, is able to accurately track italic and non-italic portions of documents. Richer models complicate inference so we present a new, streamlined procedure that is over 25x faster than the method used by Berg-Kirkpatrick et al. (2013). Our final system achieves a relative word error reduction of 22% compared to state-of-the-art results on a dataset of historical newspapers.

Robust Logistic Regression using Shift Parameters
Julie Tibshirani and Christopher Manning

Annotation errors can significantly hurt classifier performance, yet datasets are only growing noisier with the increased use of Amazon Mechanical Turk and techniques like distant supervision that automatically generate labels. In this paper, we present a robust extension of logistic regression that incorporates the possibility of mislabelling directly into the objective. This model can be trained through nearly the same means as logistic regression, and retains its efficiency on high-dimensional datasets. We conduct initial experiments on named entity recognition data and find that our approach can provide a significant improvement over the standard model when annotation errors are present.

Faster Phrase-Based Decoding by Refining Feature State
Kenneth Heafield, Michael Kayser, and Christopher Manning

We contribute a faster decoding algorithm for phrase-based machine translation. Translation hypotheses keep track of state, such as context for the language model and coverage of words in the source sentence. Most features depend upon only part of the state, but traditional algorithms, including cube pruning, handle state atomically. For example, cube pruning will repeatedly query the language model with hypotheses that differ only in source coverage, despite the fact that source coverage is irrelevant to the language model. Our key contribution avoids this behavior by placing hypotheses into equivalence classes, masking the parts of state that matter least to the score. Moreover, we exploit shared words in hypotheses to iteratively refine language model scores rather than handling language model state atomically. Since our algorithm and cube pruning are both approximate, improvement can be used to increase speed or accuracy. When tuned to attain the same accuracy, our algorithm is 4.0-7.7 times as fast as the Moses decoder with cube pruning.

Decoder Integration and Expected BLEU Training for Recurrent Neural Network Language Models
Michael Auli and Jianfeng Gao

Neural network language models are often trained by optimizing likelihood, but we would prefer to optimize for a task specific metric, such as BLEU in machine translation. We show how a recurrent neural network language model can be optimized towards an expected BLEU loss instead of the usual cross-entropy criterion. Furthermore, we tackle the issue of directly integrating a recurrent network into first-pass decoding under an efficient approximation. Our best results improve a phrase-based statistical machine translation system trained on WMT 2012 French-English data by up to 2.0 BLEU, and the expected BLEU objective improves over a cross-entropy trained model by up to 0.6 BLEU in a single reference setup.

On the Elements of an Accurate Tree-to-String Machine Translation System
Graham Neubig and Kevin Duh

While tree-to-string (T2S) translation theoretically holds promise for efficient, accurate translation, in previous reports T2S systems have often proven inferior to other machine translation (MT) methods such as phrase-based or hierarchical phrase-based MT. In this paper, we attempt to clarify the reason for this performance gap by investigating a number of peripheral elements that affect the accuracy of T2S systems, including parsing, alignment, and search. Based on detailed experiments on the English-Japanese and Japanese-English pairs, we show how a basic T2S system that performs on par with phrase-based systems can be improved by 2.6-4.6 BLEU, greatly exceeding existing state-of-the-art methods. These results indicate that T2S systems indeed hold much promise, but the above-mentioned elements must be taken seriously in construction of these systems.

Simple extensions and POS Tags for a reparameterised IBM Model 2
Douwe Gelling and Trevor Cohn

A modification of a reparameterisation of IBM Model 2 is presented, which makes the model more flexible, and able to model a preference for aligning to words to either the right or left, and take into account POS tags on the target side of the corpus. We show that this extension has a very small impact on training times, while obtain- ing better alignments in terms of BLEU scores.

Dependency-based Pre-ordering for Chinese-English Machine Translation
Jingsheng Cai, Masao Utiyama, Eiichiro Sumita, and Yujie Zhang

In statistical machine translation (SMT), syntax-based pre-ordering of the source language is an effective method for dealing with language pairs where there are great differences in their respective word orders. This paper introduces a novel pre-ordering approach based on dependency parsing for Chinese-English SMT. We present a set of dependency-based pre-ordering rules which improved the BLEU score by 1.61 on the NIST 2006 evaluation data. We also investigate the accuracy of the rule set by conducting human evaluations.

Generalized Character-Level Spelling Error Correction
Noura Farra, Nadi Tomeh, Alla Rozovskaya, and Nizar Habash

We present a generalized discriminative model for spelling error correction which targets character-level transformations. While operating at the character level, the model makes use of word-level and contextual information. In contrast to previous work, the proposed approach learns to correct a variety of error types without guidance of manually-selected constraints or language-specific features. We apply the model to correct errors in Egyptian Arabic dialect text, achieving 65

Improved Iterative Correction for Distant Spelling Errors
Sergey Gubanov, Irina Galinskaya, and Alexey Baytin

Noisy channel models, widely used in modern spellers, cope with typical misspellings,but do not work well with infrequent and difficult spelling errors. In this paper, we have improved the noisy channel approach by iterative stochastic search for the best correction. The proposed algorithm allowed us to avoid local minima problem and improve the F1 measure by 6.6% on distant spelling errors.

Predicting Grammaticality on an Ordinal Scale
Michael Heilman, Aoife Cahill, Nitin Madnani, Melissa Lopez, Matthew Mulholland, and Joel Tetreault

Automated methods for identifying whether sentences are grammatical have various potential applications (e.g., machine translation, automated essay scoring, computer-assisted language learning). In this work, we construct a statistical model of grammaticality using various linguistic features (e.g., misspelling counts, parser outputs, n-gram language model scores). We also present a new publicly available dataset of learner sentences judged for grammaticality on an ordinal scale. In evaluations, we compare our system to the one from Post (2011) and find that our approach yields state-of-the-art performance.

I'm a Belieber: Social Roles via Self-identification and Conceptual Attributes
Charley Beller, Rebecca Knowles, Craig Harman, Shane Bergsma, Margaret Mitchell, and Benjamin Van Durme

Motivated by work predicting coarse-grained author categories in social media, such as gender or political preference, we explore whether Twitter contains information to support the prediction of fine-grained categories, or social roles. We find that the simple self-identification pattern "I am a ___" supports significantly richer classification than previously explored, successfully retrieving a variety of fine-grained roles. For a given role (e.g., writer), we can further identify characteristic attributes using a simple possessive construction (e.g., "writer's ___"). Tweets that incorporate the attribute terms in first person possessives ("my ___") are confirmed to be an indicator that the author holds the associated social role.

Automatically Detecting Corresponding Edit-Turn-Pairs in Wikipedia
Johannes Daxenberger and Iryna Gurevych

In this study, we analyze links between edits in Wikipedia articles and turns from their discussion page. Our motivation is to better understand implicit details about the writing process and knowledge flow in collaboratively created resources. Based on properties of the involved edit and turn, we have defined constraints for corresponding edit-turn-pairs. We manually annotated a corpus of 636 corresponding and non-corresponding edit-turn-pairs. Furthermore, we show how our data can be used to automatically identify corresponding edit-turn-pairs. With the help of supervised machine learning, we achieve an accuracy of .87 for this task.

Two Knives Cut Better Than One: Chinese Word Segmentation with Dual Decomposition
Mengqiu Wang, Rob Voigt, and Christopher Manning

There are two dominant approaches to Chinese word segmentation: word-based and character-based models, each with respective strengths. Prior work has shown that gains in segmentation performance can be achieved from combining these two types of models; however, past efforts have not provided a practical technique to allow mainstream adoption. We propose a method that effectively combines the strength of both segmentation schemes using an efficient dual-decomposition algorithm for joint inference. Our method is simple and easy to implement. Experiments on SIGHAN 2003 and 2005 evaluation datasets show that our method achieves the best reported results to date on 6 out of 7 datasets.

Effective Document-Level Features for Chinese Patent Word Segmentation
Si Li and Nianwen Xue

A patent is a property right for an invention granted by the government to the inventor. Patents often have a high concentration of scientific and technical terms that are rare in everyday language. However, some scientific and technical terms usually appear with high frequency only in one specific patent. In this paper, we propose a pragmatic approach to Chinese word segmentation on patents where we train a sequence labeling model based on a group of novel document-level features. Experiments show that the accuracy of our model reached 96.3% (F1 score) on development set and 95.0% on a held-out test set.

Word Segmentation of Informal Arabic with Domain Adaptation
Will Monroe, Spence Green, and Christopher Manning

Segmentation of clitics has been shown to improve accuracy on a variety of Arabic NLP tasks. However, state-of-the-art Arabic word segmenters are either limited to formal Modern Standard Arabic, performing poorly on Arabic text featuring dialectal vocabulary and grammar, or rely on linguistic knowledge that is hand-tuned for each dialect. We extend an existing MSA segmenter with a simple domain adaptation technique and new features in order to segment informal and dialectal Arabic text. Experiments show that our system outperforms existing systems on newswire, broadcast news and Egyptian dialect, improving segmentation F1 score on a recently released Egyptian Arabic corpus to 95.1%, compared to 90.8% for another segmenter designed specifically for Egyptian Arabic.

Resolving Lexical Ambiguity in Tensor Regression Models of Meaning
Dimitri Kartsaklis, Nal Kalchbrenner, and Mehrnoosh Sadrzadeh

This paper provides a method for improving tensor-based compositional distributional models of meaning by the addition of an explicit disambiguation step prior to composition. In contrast with previous research where this hypothesis has been successfully tested against relatively simple compositional models, in our work we use a robust model trained with linear regression. The results we get in two experiments show the superiority of the prior disambiguation method and suggest that the effectiveness of this approach is model-independent.

A Novel Content Enriching Model for Microblog Using News Corpus
Yunlun Yang, Zhihong Deng, and Hongliang Yu

In this paper, we propose a novel model for enriching the content of microblogs by exploiting external knowledge, thus improving the data sparseness problem in short text classification. We assume that microblogs share the same topics with external knowledge. We first build an optimization model to infer the topics of microblogs by employing the topic-word distribution of the external knowledge. Then the content of microblogs is further enriched by relevant words from external knowledge. Experiments on microblog classification show that our approach is effective and outperforms traditional text classification methods.

Learning Bilingual Word Representations by Marginalizing Alignments
Tomáš Kočiský, Karl Moritz Hermann, and Phil Blunsom

We present a probabilistic model that simultaneously learns alignments and distributed representations for bilingual data. By marginalizing over word alignments the model captures a larger semantic context than prior work relying on hard alignments. The advantage of this approach is demonstrated in a cross-lingual classification task, where we outperform the prior published state of the art.

Detecting Retries of Voice Search Queries
Rivka Levitan and David Elson

When a system fails to correctly recognize a voice search query, the user will frequently retry the query, either by repeating it exactly or rephrasing it in an attempt to adapt to the system's failure. It is desirable to be able to identify queries as retries both offline, as a valuable quality signal, and online, as contextual information that can aid recognition. We present a method than can identify retries offline with 81

Sliding Alignment Windows for Real-Time Crowd Captioning
Mohammad Kazemi, Rahman Lavaee, Iftekhar Naim, and Daniel Gildea

The primary way of providing real-time speech to text captioning for hard of hearing people is to employ expensive professional stenographers who can type as fast as natural speaking rates. Recent work has shown that a feasible alternative is to combine the partial captions of ordinary typists, each of whom is able to type only part of what they hear. In this paper, we extend the state of the art fixed-window alignment algorithm (Naim et al., 2013) for combining the individual captions into a final output sequence. Our method performs alignment on a sliding window of the input sequences, drastically reducing both the number of errors and the latency of the system to the end user over the previously published approaches.

Detection of Topic and its Extrinsic Evaluation Through Multi-Document Summarization
Yoshimi Suzuki and Fumiyo Fukumoto

This paper presents a method for detecting words related to a topic (we call them topic words) over time in the stream of documents. Topic words are widely distributed in the stream of documents, and sometimes they frequently appear in the documents, and sometimes not. We propose a method to reinforce topic words with

low frequencies by collecting documents from the corpus, and applied Latent Dirichlet Allocation to these documents. For the results of LDA, we identified topic words by using Moving Average Convergence Divergence. In order to evaluate the method, we applied the results of topic detection to extractive multi-document summarization. The results showed that the method was effective for sentence selection in summarization.

Content Importance Models for Scoring Writing From Sources
Beata Beigman Klebanov, Nitin Madnani, Jill Burstein, and Swapna Somasundaran

Selection of information from external sources is an important skill assessed in educational measurement. We address an integrative summarization task used in an assessment of English proficiency for non-native speakers applying to higher education institutions in the USA. We evaluate a variety of content importance models that help predict which parts of the source material should be selected by the test-taker in order to succeed on this task.

Chinese Morphological Analysis with Character-level POS Tagging
Mo Shen, Hongxiao Liu, Daisuke Kawahara, and Sadao Kurohashi

The focus of recent studies on Chinese word segmentation, part-of-speech (POS) tagging and parsing has been shifting from words to characters. However, existing methods have not yet fully utilized the potentials of Chinese characters. In this paper, we investigate the usefulness of character-level part-of-speech in the task of Chinese morphological analysis. We propose the first tagset designed for the task of character-level POS tagging. We propose a method that performs character-level POS tagging jointly with word segmentation and word-level POS tagging. Through experiments, we demonstrate that by introducing character-level POS information, the performance of a baseline morphological analyzer can be significantly improved.

Part-of-Speech Tagging using Conditional Random Fields: Exploiting Sub-Label Dependencies for Improved Accuracy
Miikka Silfverberg, Teemu Ruokolainen, Krister Lindén, and Mikko Kurimo

We discuss part-of-speech (POS) tagging in presence of large, fine-grained label sets using conditional random fields (CRFs). We propose improving tagging accuracy by utilizing dependencies within sub-components of the fine-grained labels. These sub-label dependencies are incorporated into the CRF model via a (relatively) straightforward feature extraction scheme. Experiments on five languages show that the approach can yield significant improvement in tagging accuracy in case the labels have sufficiently rich inner structure.

POS induction with distributional and morphological information using a distance-dependent Chinese restaurant process
Kairit Sirts, Jacob Eisenstein, Micha Elsner, and Sharon Goldwater

We present a new approach to inducing the syntactic categories of words, combining their distributional and morphological properties in a joint nonparametric Bayesian model based on the distance-dependent Chinese Restaurant Process. In this framework, a log-linear model of morphological similarity acts as a prior distribution over word clusterings; the likelihood function is the probability of generating vector word embeddings. The weights of the morphology model are learned jointly while inducing part-of-speech clusters, and are therefore driven to cohere with the distributional features. The resulting algorithm outperforms competitive alternatives on English POS induction.

Improving the Recognizability of Syntactic Relations Using Contextualized Examples
Aditi Muralidharan and Marti Hearst

A common task in qualitative data analysis is to characterize the usage of a linguistic entity by issuing queries over syntactic relations between words. Previous interfaces for searching over syntactic structures require programming-style queries. User interface research suggests that it is easier to recognize a pattern than to compose it from scratch; therefore, interfaces for non-experts should show previews of syntactic relations. What these previews should look like is an open question that we explored with a 400-participant Mechanical Turk experiment. We found that syntactic relations are recognized with 34% higher accuracy when contextual examples are shown than a baseline of naming the relations alone. This suggests that user interfaces should display contextual examples of syntactic relations to help users choose between different relations.

Posters: SRW papers

The SRW poster session includes the four thesis proposals presented in Session 2D (p.32) and nine papers presented in the SRW oral session (p.41).

Bayesian Kernel Methods for Natural Language Processing
Daniel Beck

Extracting Temporal and Causal Relations between Events
Paramita Mirza

Towards a discourse relation-aware approach for Chinese-English machine translation
Frances Yung

Analyzing Positions and Topics in Political Discussions of the German Bundestag
Cäcilia Zirn

A Mapping-Based Approach for General Formal Human Computer Interaction Using Natural Language
Vincent Letard, Sophie Rosset, and Gabriel Illouz

An Exploration of Embeddings for Generalized Phrases
Wenpeng Yin and Hinrich Schütze

Learning Grammar with Explicit Annotations for Subordinating Conjunctions
Dongchen Li, Xiantao Zhang, and Xihong Wu

Going beyond sentences when applying tree kernels
Dmitry Ilvovsky

Multi-document summarization using distortion-rate ratio
Ulukbek Attokurov and Ulug Bayazit

Disambiguating prepositional phrase attachment sites with sense information captured in contextualized distributional data
Clayton Greenberg

Open Information Extraction for Spanish Language based on Syntactic Constraints
Alisa Zhila and Alexander Gelbukh

Improving Text Normalization via Unsupervised Model and Discriminative Reranking
Chen Li and Yang Liu

Semi-Automatic Development of KurdNet, The Kurdish WordNet
Purya Aliabadi

Demonstrations

Cross-Lingual Information to the Rescue in Keyword Extraction
Chung-Chi Huang, Maxine Eskenazi, Jaime Carbonell, Lun-Wei Ku, and Ping-Che Yang
We introduce a method that extracts keywords in a language with the help of the other. In our approach, we bridge and fuse conventionally irrelevant word statistics in languages. The method involves estimating preferences for keywords w.r.t. domain topics and generating cross-lingual bridges for word statistics integration. At run-time, we transform parallel articles into word graphs, build cross-lingual edges, and exploit PageRank with word keyness information for keyword extraction. We present the system, BiKEA, that applies the method to keyword analysis. Experiments show that keyword extraction benefits from PageRank, globally

learned keyword preferences, and cross-lingual word statistics interaction which respects language diversity.

Visualization, Search, and Error Analysis for Coreference Annotations
Markus Gärtner, Anders Björkelund, Gregor Thiele, Wolfgang Seeker, and Jonas Kuhn

We present the ICARUS Coreference Explorer, an interactive tool to browse and search coreference-annotated data. It can display coreference annotations as a tree, as an entity grid, or in a standard text-based display mode, and lets the user switch freely between the different modes. The tool can compare two different annotations on the same document, allowing system developers to evaluate errors in automatic system predictions. It features a flexible search engine, which enables the user to graphically construct search queries over sets of documents annotated with coreference.

Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition
Jana Straková, Milan Straka, and Jan Hajič

We present two recently released open-source taggers: NameTag is a free software for named entity recognition (NER) which achieves state-of-the-art performance on Czech; MorphoDiTa (Morphological Dictionary and Tagger) performs morphological analysis (with lemmatization), morphological generation, tagging and tokenization with state-of-the-art results for Czech and a throughput around 10-200K words per second. The taggers can be trained for any language for which annotated data exist, but they are specifically designed to be efficient for inflective languages, Both tools are free software under LGPL license and are distributed along with trained linguistic models which are free for non-commercial use under the CC~BY-NC-SA license. The releases include standalone tools, C++ libraries with Java, Python and Perl bindings and web services.

Community Evaluation and Exchange of Word Vectors at wordvectors.org
Manaal Faruqui and Chris Dyer

Vector space word representations are useful for many natural language processing applications. The diversity of techniques for computing vector representations and the large number of evaluation benchmarks makes reliable comparison a tedious task both for researchers developing new vector space models and for those wishing to use them. We present a website and suite of offline tools that that facilitate evaluation of word vectors on standard lexical semantics benchmarks and permit exchange and archival by users who wish to find good vectors for their applications. The system is accessible at: www.wordvectors.org.

WINGS:Writing with Intelligent Guidance and Suggestions
Xianjun Dai, Yuanchao Liu, Xiaolong Wang, and Bingquan Liu

Without inspirations, writing may be a frustrating task for most people. In this study, we designed and implemented WINGS, a Chinese input method extended on IBus-Pinyin with intelligent writing assistance. In addition to supporting common Chinese input, WINGS mainly attempts to spark users' inspirations by recommending both word level and sentence level writing suggestions. The main strategies used by WINGS, including providing syntactically and semantically related words based on word vector representation and recommending contextually related sentences based on LDA, are discussed and described. Experimental results suggest that WINGS can facilitate Chinese writing in an effective and creative manner.

DKPro Keyphrases: Flexible and Reusable Keyphrase Extraction Experiments
Nicolai Erbs, Pedro Bispo Santos, Iryna Gurevych, and Torsten Zesch

DKPro Keyphrases is a keyphrase extraction framework based on UIMA. It offers a wide range of state-of-the-art keyphrase experiments approaches. At the same time, it is a workbench for developing new extraction approaches and evaluating their impact. DKPro Keyphrases is publicly available under an open-source license.

Real-Time Detection, Tracking, and Monitoring of Automatically Discovered Events in Social Media
Miles Osborne, Sean Moran, Richard McCreadie, Alexander Von Lunen, Martin Sykora, Elizabeth Cano, Neil Ireson, Craig Macdonald, Iadh Ounis, Yulan He, Tom Jackson, Fabio Ciravegna, and Ann O'Brien

We introduce *ReDites*, a system for real-time event detection, tracking, monitoring and visualisation. It is designed to assist Information Analysts in understanding and exploring complex events as they unfold in the world. Events are automatically detected from the Twitter stream. Then those that are categorised as being security-relevant are tracked, geolocated, summarised and visualised for the end-user. Furthermore, the system tracks changes in emotions over events, signalling possible flashpoints or abatement. We demonstrate

the capabilities of *ReDites* using an extended use case from the September 2013 Westgate shooting incident. Through an evaluation of system latencies, we also show that enriched events are made available for users to explore within seconds of that event occurring.

The Excitement Open Platform for Textual Inferences
Bernardo Magnini, Roberto Zanoli, Ido Dagan, Kathrin Eichler, Guenter Neumann, Tae-Gil Noh, Sebastian Padó, Asher Stern, and Omer Levy

This paper presents the Excitement Open Platform (EOP), a generic architecture and a comprehensive implementation for textual inference in multiple languages. The platform includes state-of-art algorithms, a large number of knowledge resources, and facilities for experimenting and testing innovative approaches. The EOP is distributed as an open source software.

WELT: Using Graphics Generation in Linguistic Fieldwork
Morgan Ulinski, Anusha Balakrishnan, Bob Coyne, Julia Hirschberg, and Owen Rambow

We describe the WordsEye Linguistics tool (WELT), a novel tool for the documentation and preservation of endangered languages. WELT is based on WordsEye (Coyne and Sproat, 2001), a text-to-scene tool that automatically generates 3D scenes from written input. WELT has two modes of operation. In the first mode, English input automatically generates a picture which can be used to elicit a description in the target language. In the second mode, the linguist formally documents the grammar of an endangered language, thereby creating a system that takes input in the endangered language and generates a picture according to the grammar; the picture can then be used to verify the grammar with native speakers. We will demonstrate WELT's use on scenarios involving Arrernte and Nahuatl.

The Stanford CoreNLP Natural Language Processing Toolkit
Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky

We describe the design and use of the Stanford CoreNLP toolkit, an extensible pipeline that provides core natural language analysis. This toolkit is quite widely used, both in the research NLP community and also among commercial and government users of open source NLP technology. We suggest that this follows from a simple, approachable design, straightforward interfaces, the inclusion of robust and good quality analysis components, and not requiring use of a large amount of associated baggage.

4

Main Conference: Tuesday, June 24

Overview

7:30 – 6:00	**Registration**				(Grand Foyer)
7:30 – 9:00	**Breakfast**				(Grand Foyer)
9:00 – 10:00	**Keynote Address: Zoran Popovic**				(Grand Ballroom V–VI)
10:00 – 10:30	**Coffee break**				(Grand Foyer)
	Session 4				
10:30 – 12:10	Machine Learning for NLP	Information Extraction II	Machine Translation II	Summarization	Language Resources and Evaluation II
12:10 – 1:30	**Lunch break**				
	Session 5				
1:30 – 2:45	Question Answering	Information Extraction III	Lexical Sematics and Ontology I	Syntax, Parsing and Tagging IV	Cognitive Modeling and Psycholinguistics
2:45 – 3:15	**Coffee break**				(Grand Foyer)
	Session 6				
3:15 – 4:30	Machine Translation III	Lexical Semantics and Ontology II	Generation / Summarization / Dialogue	NLP Applications and NLP Enabled Technology I	Language Resources and Evaluation III
4:50 – 7:20	**Poster and Dinner Session II: Long Papers, Short Papers, and Demonstrations**				(Grand Ballroom I–X)
7:30 – 10:00	**Social at the National Aquarium in Baltimore**				

61

Keynote Address: Zoran Popović

"Text Generation for Infinitely Adaptable Curricula"

Tuesday, June 24, 2014, 9:00–10:00am

Grand Ballroom V–VI

Abstract: Recent studies show that to achieve mastery of a topic by 95% of the student population, some students need ten times more learning content than is available in current curricula. At issue is not just increased volume, but the need for a highly differentiated content specialized to promote optimal learning for each unique learner. To address this synthesis problem we have developed a generative platform capable of dynamically varying content based on the individual student needs. This approach recently achieved 93% mastery of a key algebra concept even for primary school students in three state-wide challenges. In this talk I will describe our work on extending the platform to enable students to solve all word problems in high-school within their preferred context (e.g. sci-fi, medieval, Harry Potter), as well as to automatically generate adaptive learning progressions for reading comprehension curricula in middle school.

Biography: Zoran Popović is the Director of the Center for Game Science at the University of Washington and the founder of Engaged Learning. Trained as a computer scientist, his research focus is on creating interactive engaging environments for learning and scientific discovery. His laboratory created Foldit, a biochemistry game that produced three Nature publications in just two years, and award-winning math learning games played by over five million learners worldwide. He is currently focusing on engaging methods that can rapidly develop experts in arbitrary domains with particular focus on revolutionizing K–12 math education. He has recently founded Engaged Learning to apply his work on generative adaptation to any curricula towards the goal of achieving school mastery by 95% of students. His contributions to the field of interactive computer graphics have been recognized by a number of awards including the NSF CAREER Award, Alfred P. Sloan Fellowship and ACM SIGGRAPH Significant New Researcher Award.

Parallel Session 4 Overview - Tuesday, June 24, 2014

Track A	Track B	Track C	Track D	Track E	
Machine Learning for NLP	*Information Extraction II*	*Machine Translation II*	*Summarization*	*Language Resources and Evaluation II*	
Grand V	Grand VI	Harborside A/B	Harborside C	Harborside D/E	
Kneser-Ney Smoothing on Expected Counts *Zhang and Chiang*	Robust Domain Adaptation for Relation Extraction via Clustering Consistency *Nguyen, Tsang, Chai, and Chieu*	Enhancing Grammatical Cohesion: Generating Transitional Expressions for SMT *Tu, Zhou, and Zong*	Modelling Events through Memory-based, Open-IE Patterns for Abstractive Summarization *Pighin, Cornolti, Alfonseca, and Filippova*	[TACL] Segmentation for Efficient Supervised Language Annotation with an Explicit Cost-Utility Tradeoff *Sperber, Simantzik, Neubig, Nakamura, and Waibel*	10:30
Robust Entity Clustering via Phylogenetic Inference *Andrews, Eisner, and Dredze*	Encoding Relation Requirements for Relation Extraction via Joint Inference *Chen, Feng, Huang, Qin, and Zhao*	Adaptive HTER Estimation for Document-Specific MT Post-Editing *Huang, Xu, Ittycheriah, and Roukos*	Hierarchical Summarization: Scaling Up Multi-Document Summarization *Christensen, Soderland, Bansal, and Mausam*	A chance-corrected measure of inter-annotator agreement for syntax *Skjærholt*	10:55
Linguistic Structured Sparsity in Text Categorization *Yogatama and Smith*	Medical Relation Extraction with Manifold Models *Wang and Fan*	Translation Assistance by Translation of L1 Fragments in an L2 Context *Gompel and Bosch*	Query-Chain Focused Summarization *Baumel, Cohen, and Elhadad*	Two Is Bigger (and Better) Than One: the Wikipedia Bitaxonomy Project *Flati, Vannella, Pasini, and Navigli*	11:20
Perplexity on Reduced Corpora *Kobayashi*	Distant Supervision for Relation Extraction with Matrix Completion *Fan, Zhao, Zhou, Liu, Zheng, and Chang*	Response-based Learning for Grounded Machine Translation *Riezler, Simianer, and Haas*	Exploiting Timelines to Enhance Multi-document Summarization *Ng, Chen, Kan, and Li*	[TACL] The Language Demographics of Amazon Mechanical Turk *Pavlick, Post, Irvine, Kachaev, and Callison-Burch*	11:45

Parallel Session 4

Session 4A: Machine Learning for NLP
Grand V   *Chair: Kevin Duh*

Kneser-Ney Smoothing on Expected Counts
Hui Zhang and David Chiang   10:30–10:55

Widely used in speech and language processing, Kneser-Ney (KN) smoothing has consistently been shown to be one of the best-performing smoothing methods. However, KN smoothing assumes integer counts, limiting its potential uses—for example, inside Expectation-Maximization. In this paper, we propose a generalization of KN smoothing that operates on fractional counts, or, more precisely, on *distributions* over counts. We rederive all the steps of KN smoothing to operate on count distributions instead of integral counts, and apply it to two tasks where KN smoothing was not applicable before: one in language model adaptation, and the other in word alignment. In both cases, our method improves performance significantly.

Robust Entity Clustering via Phylogenetic Inference
Nicholas Andrews, Jason Eisner, and Mark Dredze   10:55–11:20

Entity clustering must determine when two named-entity mentions refer to the same entity. Typical approaches use a pipeline architecture that clusters the mentions using fixed or learned measures of name and context similarity. In this paper, we propose a model for cross-document coreference resolution that achieves robustness by learning similarity from unlabeled data. The generative process assumes that each entity mention arises from copying and optionally mutating an earlier name from a similar context. Clustering the mentions into entities depends on recovering this copying tree jointly with estimating models of the mutation process and parent selection process. We present a block Gibbs sampler for posterior inference and an empirical evaluation on several datasets.

Linguistic Structured Sparsity in Text Categorization
Dani Yogatama and Noah A. Smith   11:20–11:45

We introduce three linguistically motivated structured regularizers based on parse trees, topics, and hierarchical word clusters for text categorization. These regularizers impose linguistic bias in feature weights, enabling us to incorporate prior knowledge into conventional bag-of-words models. We show that our structured regularizers consistently improve classification accuracies compared to standard regularizers that penalize features in isolation (such as lasso, ridge, and elastic net regularizers) on a range of datasets for various text prediction problems: topic classification, sentiment analysis, and forecasting.

Perplexity on Reduced Corpora
Hayato Kobayashi   11:45–12:10

This paper studies the idea of removing low-frequency words from a corpus, which is a common way to reduce computational costs, from a theoretical standpoint. Based on the assumption that a corpus follows Zipf's law, we derive trade-off formulae of the perplexity of k-gram models and topic models with respect to the size of the reduced vocabulary. In addition, we show the approximate behavior of each formula under certain conditions. We verify the correctness of our theory on synthetic corpora and examine the gap between theory and practice on real corpora.

Session 4B: Information Extraction II
Grand VI   *Chair: Jian Su*

Robust Domain Adaptation for Relation Extraction via Clustering Consistency
Minh Luan Nguyen, Ivor W. Tsang, Kian Ming A. Chai, and Hai Leong Chieu   10:30–10:55

We propose a two-phase framework to adapt existing relation extraction classifiers to extract relations for new target domains. We address two challenges: negative transfer when knowledge in source domains is used without considering the differences in relation distributions; and lack of adequate labeled samples for rarer relations in the new domain, due to a small labeled data set and imbalance relation distributions. Our framework leverages on both labeled and unlabeled data in the target domain. First, we determine the relevance of each source domain to the target domain for each relation type, using the consistency between the clustering given by the target domain labels and the clustering given by the predictors trained for the source domain. To overcome the lack of labeled samples for rarer relations, these clusterings operate on both the labeled and unlabeled data in the target domain. Second, we trade-off between using relevance-weighted source-domain predictors and the labeled target data. Again, to overcome the imbalance distribution, the source-domain predictors operate on the unlabeled target data. Our method outperforms numerous baselines and a weakly-supervised relation extraction method on ACE 2004 and YAGO.

Encoding Relation Requirements for Relation Extraction via Joint Inference
Liwei Chen, Yansong Feng, Songfang Huang, Yong Qin, and Dongyan Zhao   10:55–11:20

Most existing relation extraction models make predictions for each entity pair locally and individually, while ignoring implicit global clues available in the knowledge base, sometimes leading to conflicts among single predictions from different entity pairs. In this paper, we propose a joint inference framework that utilizes these global clues to resolve disagreements among local predictions. We exploit two kinds of clues to generate constraints to capture the implicit type and cardinality requirements of a relation. Experimental results on three datasets, in both English and Chinese, show that our framework outperforms the state-of-the-art relation exraction models when such clues are applicable to the datasets. And, we find that the clues learned automatically from existing knowledge bases perform comparably to those refined by human.

Medical Relation Extraction with Manifold Models
Chang Wang and James Fan   11:20–11:45

In this paper, we present a manifold model for medical relation extraction. Our model is built upon a medical corpus containing 80M sentences (11 gigabyte text) and designed to accurately and efficiently detect the key medical relations that can facilitate clinical decision making. Our approach integrates domain specific parsing and typing systems, and can utilize labeled as well as unlabeled examples. To provide users with more flexibility, we also take label weight into consideration. Effectiveness of our model is demonstrated both theoretically with a proof to show that the solution is a closed-form solution and experimentally with positive results in experiments.

Distant Supervision for Relation Extraction with Matrix Completion
Miao Fan, Deli Zhao, Qiang Zhou, Zhiyuan Liu, Thomas Fang Zheng, and Edward Y. Chang
11:45–12:10

The essence of distantly supervised relation extraction is that it is an incomplete multi-label classification problem with sparse and noisy features. To tackle the sparsity and noise challenges, we propose solving the classification problem using matrix completion on factorized matrix of minimized rank. We formulate relation classification as completing the unknown labels of testing items (entity pairs) in a sparse matrix that concatenates training and testing textual features with training labels. Our algorithmic framework is based on the assumption that the rank of item-by-feature and item-by-label joint matrix is low. We apply two optimization models to recover the underlying low-rank matrix leveraging the sparsity of feature-label matrix. The matrix completion problem is then solved by the fixed point continuation (FPC) algorithm, which can find the global optimum. Experiments on two widely used datasets with different dimensions of textual features demonstrate that our low-rank matrix completion approach significantly outperforms the baseline and the state-of-the-art methods.

Session 4C: Machine Translation II
Harborside A/B   *Chair: Colin Cherry*

Enhancing Grammatical Cohesion: Generating Transitional Expressions for SMT
Mei Tu, Yu Zhou, and Chengqing Zong   10:30–10:55

Transitional expressions provide glue that holds ideas together in a text and enhance the logical organization, which together help improve readability of a text. However, in most current statistical machine translation (SMT) systems, the outputs of compound-complex sentences still lack proper transitional expressions. As a result, the translations are often hard to read and understand. To address this issue, we propose two novel models to en-courage generating such transitional expressions by introducing the source compound-complex sentence structure (CSS). Our models include a CSS-based translation model, which generates new CSS-based translation rules, and a generative transfer model, which encourages producing transitional expressions during decoding. The two models are integrated into a hierarchical phrase-based translation system to evaluate their effectiveness. The experimental results show that significant improvements are achieved on various test data while the translations are more cohesive and smooth.

Adaptive HTER Estimation for Document-Specific MT Post-Editing
Fei Huang, Jian-Ming Xu, Abraham Ittycheriah, and Salim Roukos   10:55–11:20

We present an adaptive translation quality estimation (QE) method to predict the human-targeted translation error rate (HTER) for a document-specific machine translation model. We first introduce features derived internal to the translation decoding process as well as externally from the source sentence analysis. We show the effectiveness of such features in both classification and regression of MT quality. By dynamically training the QE model for the document-specific MT model, we are able to achieve consistency and prediction quality across multiple documents, demonstrated by the higher correlation co- efficient and F-scores in finding Good sentences. Additionally, the proposed method is applied to IBM English-to-Japanese MT post editing field study and we observe strong correlation with human preference, with a 10% increase in human translators' productivity.

Translation Assistance by Translation of L1 Fragments in an L2 Context
Maarten van Gompel and Antal van den Bosch   11:20–11:45

In this paper we present new research in translation assistance. We describe a system capable of translating native language (L1) fragments to foreign language (L2) fragments in an L2 context. Practical applications of this research can be framed in the context of second language learning. The type of translation assistance system under investigation here encourages language learners to write in their target language while allowing them to fall back to their native language in case the correct word or expression is not known. These code switches are subsequently translated back to L2 given the L2 context. We study the feasibility of exploiting cross-lingual context to obtain high-quality translation suggestions that improve over statistical language mod-elling and word-sense disambiguation baselines. A classification-based approach is presented that is indeed found to improve significantly over these baselines by making use of a contextual window spanning a small number of neighbouring words.

Response-based Learning for Grounded Machine Translation
Stefan Riezler, Patrick Simianer, and Carolin Haas   11:45–12:10

We propose a novel learning approach for statistical machine translation (SMT) that allows to extract supervision signals for structured learning from an extrinsic response to a translation input. We show how to generate responses by grounding SMT in the task of executing a semantic parse of a translated query against a database. Experiments on the Geoquery database show an improvement of about 6 points in F1-score for response-based learning over learning from references only on returning the correct answer from a semantic parse of a translated query. In general, our approach alleviates the dependency on human reference translations and solves the reachability problem in structured learning for SMT.

Session 4D: Summarization

Harborside C   *Chair: Yang Liu*

Modelling Events through Memory-based, Open-IE Patterns for Abstractive Summarization

Daniele Pighin, Marco Cornolti, Enrique Alfonseca, and Katja Filippova   10:30–10:55

Abstractive text summarization of news requires a way of representing events, such as a collection of pattern clusters in which every cluster represents an event (e.g., marriage) and every pattern in the cluster is a way of expressing the event (e.g., X married Y, X and Y tied the knot). We compare three ways of extracting event patterns: heuristics-based, compression-based and memory-based. While the former has been used previously in multi-document abstraction, the latter two have never been used for this task. Compared with the first two techniques, the memory-based method allows for generating significantly more grammatical and informative sentences, at the cost of searching a vast space of hundreds of millions of parse trees of known grammatical utterances. To this end, we introduce a data structure and a search method that make it possible to efficiently extrapolate from every sentence the parse sub-trees that match against any of the stored utterances.

Hierarchical Summarization: Scaling Up Multi-Document Summarization

Janara Christensen, Stephen Soderland, Gagan Bansal, and Mausam   10:55–11:20

Current multi-document summarization (MDS) systems have been designed for short, unstructured summaries of 10-15 documents, and are inadequate for larger document collections. We propose a new approach to scaling up summarization called hierarchical summarization, and present the first implemented system, Summa. Summa produces a hierarchy of relatively short summaries, where the top level provides a general overview and users can navigate the hierarchy to drill down for more details on topics of interest. Summa optimizes for coherence as well as coverage of salient information. In an Amazon Mechanical Turk evaluation, users prefer Summa ten times as often as flat MDS and three times as often as timelines.

Query-Chain Focused Summarization

Tal Baumel, Raphael Cohen, and Michael Elhadad   11:20–11:45

Update summarization is a form of multi-document summarization where a document set must be summarized in the context of other documents assumed to be known. Efficient update summarization must focus on identifying new information and avoiding repetition of known information. In Query-focused summarization, the task is to produce a summary as an answer to a given query. We introduce a new task, Query-Chain Summarization, which combines aspects of the two previous tasks: starting from a given document set, increasingly specific queries are consid-ered, and a new summary is produced at each step. This process models exploratory search: a user explores a new topic by submitting a sequence of queries, inspecting a summary of the result set and phrasing a new query at each step. We present a novel dataset comprising 22 query-chains sessions of length 3 with 3 matching human summaries each in the consumer-health domain. Our analysis demonstrates that summaries produced in the context of such exploratory process are different from informative summaries. We present an algorithm for Query-Chain Summarization based on a new LDA topic model variant. Evaluation indicates the algorithm improves on strong baselines.

Exploiting Timelines to Enhance Multi-document Summarization

Jun-Ping Ng, Yan Chen, Min-Yen Kan, and Zhoujun Li   11:45–12:10

We study the use of temporal information in the form of timelines to enhance multi-document summarization. We employ a fully automated temporal processing system to generate a timeline for each input document. We derive three features from these timelines, and show that their use in supervised summarization lead to a significant 4.1% improvement in ROUGE performance over a state-of-the-art baseline. In addition, we propose TIMEMMR, a modification to Maximal Marginal Relevance that promotes temporal diversity by way of computing time span similarity, and show its utility in summarizing certain document sets. We also propose a filtering metric to discard noisy timelines generated by our automatic processes, to purify the timeline input for summarization. By selectively using timelines guided by filtering, overall summarization performance is increased by a significant 5.9%.

Session 4E: Language Resources and Evaluation II
Harborside D/E   *Chair: Nianwen Xue*

[TACL] Segmentation for Efficient Supervised Language Annotation with an Explicit Cost-Utility Tradeoff
Matthias Sperber, Mirjam Simantzik, Graham Neubig, Satoshi Nakamura, and Alex Waibel
10:30–10:55

In this paper, we study the problem of manually correcting automatic annotations of natural language in as efficient a manner as possible. We introduce a method for automatically segmenting a corpus into chunks such that many uncertain labels are grouped into the same chunk, while human supervision can be omitted altogether for other segments. A tradeoff must be found for segment sizes. Choosing short segments allows us to reduce the number of highly confident labels that are supervised by the annotator, which is useful because these labels are often already correct and supervising correct labels is a waste of effort. In contrast, long segments reduce the cognitive effort due to context switches. Our method helps find the segmentation that optimizes supervision efficiency by defining user models to predict the cost and utility of supervising each segment and solving a constrained optimization problem balancing these contradictory objectives. A user study demonstrates substantial gains over pre-segmented, confidence-ordered baselines on two natural language processing tasks: speech transcription and word segmentation.

A chance-corrected measure of inter-annotator agreement for syntax
Arne Skjærholt   10:55–11:20

Following the works of Carletta (1996) and Artstein and Poesio (2008), there is an increasing consensus within the field that in order to properly gauge the reliability of an annotation effort, chance-corrected measures of inter-annotator agreement should be used. With this in mind, it is striking that virtually all evaluations of syntactic annotation efforts use uncorrected parser evaluation metrics such as bracket F1 (for phrase structure) and accuracy scores (for dependencies). In this work we present a chance-corrected metric based on Krippendorff's α, adapted to the structure of syntactic annotations and applicable both to phrase structure and dependency annotation without any modifications. To evaluate our metric we first present a number of synthetic experiments to better control the sources of noise and gauge the metric's responses, before finally contrasting the behaviour of our chance-corrected metric with that of uncorrected parser evaluation metrics on real corpora.

Two Is Bigger (and Better) Than One: the Wikipedia Bitaxonomy Project
Tiziano Flati, Daniele Vannella, Tommaso Pasini, and Roberto Navigli   11:20–11:45

We present WiBi, an approach to the automatic creation of a bitaxonomy for Wikipedia, that is, an integrated taxonomy of Wikipage pages and categories. We leverage the information available in either one of the taxonomies to reinforce the creation of the other taxonomy. Our experiments show higher quality and coverage than state-of-the-art resources like DBpedia, YAGO, MENTA, WikiNet and WikiTaxonomy. WiBi is available at http://wibitaxonomy.org.

[TACL] The Language Demographics of Amazon Mechanical Turk
Ellie Pavlick, Matt Post, Ann Irvine, Dmitry Kachaev, and Chris Callison-Burch   11:45–12:10

We present a large scale study of the languages spoken by bilingual workers on Mechanical Turk (MTurk). We establish a methodology for determining the language skills of anony- mous crowd workers that is more robust than simple surveying. We validate workers' self- reported language skill claims by measuring their ability to correctly translate words, and by geolocating workers to see if they reside in countries where the languages are likely to be spoken. Rather than posting a one-off survey, we posted paid tasks consisting of 1,000 assignments to translate a total of 10,000 words in each of 100 languages. Our study ran for several months, and was highly visible on the MTurk crowdsourcing platform, increas- ing the chances that bilingual workers would complete it. Our study was useful both to cre- ate bilingual dictionaries and to act as cen- sus of the bilingual speakers on MTurk. We use this data to recommend languages with the largest speaker populations as good candidates for other researchers who want to develop crowdsourced, multilingual technologies. To further demonstrate the value of creating data via crowdsourcing, we hire workers to create bilingual parallel corpora in six Indian lan- guages, and use them to train statistical ma- chine translation systems.

Parallel Session 5 Overview - Tuesday, June 24, 2014

Track A *Question Answering* Grand V	Track B *Information Extraction III* Grand VI	Track C *Lexical Sematics and Ontology I* Harborside A/B	Track D *Syntax, Parsing and Tagging IV* Harborside C	Track E *Cognitive Modeling and Psycholinguistics* Harborside D/E	
Information Extraction over Structured Data: Question Answering with Freebase *Yao and Van Durme*	Toward Future Scenario Generation: Extracting Event Causality Exploiting Semantic Relation, Context, and Association Features *Hashimoto, Torisawa, Kloetzer, Sano, Varga, Oh, and Kidawara*	That's sick dude!: Automatic identification of word sense change across different timescales *Mitra, Mitra, Riedl, Biemann, Mukherjee, and Goyal*	A Provably Correct Learning Algorithm for Latent-Variable PCFGs *Cohen and Collins*	Weak semantic context helps phonetic learning in a model of infant language acquisition *Frank, Feldman, and Goldwater*	1:30
Knowledge-Based Question Answering as Machine Translation *Bao, Duan, Zhou, and Zhao*	Cross-narrative Temporal Ordering of Medical Events *Raghavan, Fosler-Lussier, Elhadad, and Lai*	A Step-wise Usage-based Method for Inducing Polysemy-aware Verb Classes *Kawahara, Peterson, and Palmer*	Spectral Unsupervised Parsing with Additive Tree Metrics *Parikh, Cohen, and Xing*	Bootstrapping into Filler-Gap: An Acquisition Story *Schijndel and Elsner*	1:55
Discourse Complements Lexical Semantics for Non-factoid Answer Reranking *Jansen, Surdeanu, and Clark*	Language-Aware Truth Assessment of Fact Candidates *Nakashole and Mitchell*	Structured Learning for Taxonomy Induction with Belief Propagation *Bansal, Burkett, Melo, and Klein*	[TACL] Cross-lingual Projected Expectation Regularization for Weakly Supervised Learning *Wang and Manning*	Nonparametric Learning of Phonological Constraints in Optimality Theory *Doyle, Bicknell, and Levy*	2:20

Parallel Session 5

Session 5A: Question Answering
Grand V   *Chair: Scott Yih*

Information Extraction over Structured Data: Question Answering with Freebase
Xuchen Yao and Benjamin Van Durme   1:30–1:55
Answering natural language questions using the Freebase knowledge base has recently been explored as a platform for advancing the state of the art in open domain semantic parsing. Those efforts map questions to sophisticated meaning representations that are then attempted to be matched against viable answer candidates in the knowledge base. Here we show that relatively modest information extraction techniques, when paired with a web-scale corpus, can outperform these sophisticated approaches by roughly 34% relative gain.

Knowledge-Based Question Answering as Machine Translation
Junwei Bao, Nan Duan, Ming Zhou, and Tiejun Zhao   1:55–2:20
A typical knowledge-based question answering(KB-QA) system faces two challenges: one is to transform natural language questions into their meaning representations(MRs); the other is to retrieve answers from knowledge bases (KBs) using generated MRs. Unlike previous methods which treat them in a cascaded manner, we present a translation-based approach to solve these two tasks in one unified framework. We translate questions to answers based on CYK parsing. Answers as translations of the span covered by each CYK cell are obtained by a question translation method, which first generates formal triple queries as MRs for the span based on question patterns and relation expressions, and then retrieves answers from a given KB based on triple queries generated. A linear model is defined over derivations, and minimum error rate training is used to tune feature weights based on a set of question-answer pairs. Compared to a KB-QA system using a state-of-the-art semantic parser, our method achieves better results.

Discourse Complements Lexical Semantics for Non-factoid Answer Reranking
Peter Jansen, Mihai Surdeanu, and Peter Clark   2:20–2:45
We propose a robust answer reranking model for non-factoid questions that integrates lexical semantics with discourse information, driven by two representations of discourse: a shallow representation centered around discourse markers, 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 experimentally demonstrate that the discourse structure of non-factoid answers provides information that is complementary to lexical semantic similarity between question and answer, improving performance up to 24% (relative) over a state-of-the-art model that exploits lexical semantic similarity alone. We further demonstrate excellent domain transfer of discourse information, suggesting these discourse features have general utility to non-factoid question answering.

Session 5B: Information Extraction III

Grand VI   *Chair: Wei Xu*

Toward Future Scenario Generation: Extracting Event Causality Exploiting Semantic Relation, Context, and Association Features

Chikara Hashimoto, Kentaro Torisawa, Julien Kloetzer, Motoki Sano, István Varga, Jong-Hoon Oh, and Yutaka Kidawara   1:30–1:55

We propose a supervised method of extracting event causalities like "conduct slash-and-burn agriculture" -> "exacerbate desertification" from the web using semantic relation (between nouns), context, and association features. Experiments show that our method outperforms baselines that are based on state-of-the-art methods. We also propose methods of generating future scenarios like "conduct slash-and-burn agriculture" -> "exacerbate desertification" -> "increase Asian dust (from China)" -> "asthma gets worse". Experiments show that we can generate 50,000 scenarios with 68% precision. We also generated a scenario "deforestation continues" -> "global warming worsens" -> "sea temperatures rise" -> "vibrio parahaemolyticus fouls (water)", which is written in no document in our input web corpus crawled in 2007. But the vibrio risk due to global warming was observed in Baker-Austin et al. (2013). Thus, we "predicted" the future event sequence in a sense.

Cross-narrative Temporal Ordering of Medical Events

Preethi Raghavan, Eric Fosler-Lussier, Noémie Elhadad, and Albert M. Lai   1:55–2:20

Cross-narrative temporal ordering of medical events is essential to the task of generating a comprehensive timeline over a patient's history. We address the problem of aligning multiple medical event sequences, corresponding to different clinical narratives, comparing the following approaches: (1) A novel weighted finite state transducer representation of medical event sequences that enables composition and search for decoding, and (2) Dynamic programming with iterative pairwise alignment of multiple sequences using global and local alignment algorithms. The cross-narrative coreference and temporal relation weights used in both these approaches are learned from a corpus of clinical narratives. We present results using both approaches and observe that the finite state transducer approach performs significantly better than the dynamic programming one by 6.8% for the problem of multiple-sequence alignment.

Language-Aware Truth Assessment of Fact Candidates

Ndapandula Nakashole and Tom M. Mitchell   2:20–2:45

This paper introduces FactChecker, a novel, language-aware approach to truth-finding. FactChecker differs from prior approaches in that it does not rely on iterative peer voting, instead it leverages language to infer believability of fact candidates. In particular, FactChecker makes use of linguistic features to detect if a given source objectively states facts or is speculative and opinionated. To ensure that fact candidates mentioned in similar sources have similar believability, FactChecker augments objectivity with a co-mention score to compute the overall believability score of a fact candidate. Our experiments on various datasets show that FactChecker yields higher accuracy than existing approaches.

Session 5C: Lexical Sematics and Ontology I

Harborside A/B   *Chair: Timothy Baldwin*

That's sick dude!: Automatic identification of word sense change across different timescales
Sunny Mitra, Ritwik Mitra, Martin Riedl, Chris Biemann, Animesh Mukherjee, and Pawan Goyal
1:30–1:55

In this paper, we propose an unsupervised method to identify noun sense changes based on rigorous analysis of time-varying text data available in the form of millions of digitized books. We construct distributional thesauri based networks from data at different time points and cluster each of them separately to obtain word-centric sense clusters corresponding to the different time points. Subsequently, we compare these sense clusters of two different time points to find if (i) there is birth of a new sense or (ii) if an older sense has got split into more than one sense or (iii) if a newer sense has been formed from the joining of older senses or (iv) if a particular sense has died. We conduct a thorough evaluation of the proposed methodology both manually as well as through comparison with WordNet. Manual evaluation indicates that the algorithm could correctly identify 60.4% birth cases from a set of 48 randomly picked samples and 57% split/join cases from a set of 21 randomly picked samples. Remarkably, in 44% cases the birth of a novel sense is attested by WordNet, while in 46% cases and 43% cases split and join are respectively confirmed by WordNet. Our approach can be applied for lexicography, as well as for applications like word sense disambiguation or semantic search.

A Step-wise Usage-based Method for Inducing Polysemy-aware Verb Classes
Daisuke Kawahara, Daniel W. Peterson, and Martha Palmer   1:55–2:20

We present an unsupervised method for inducing verb classes from verb uses in giga-word corpora. Our method consists of two clustering steps: verb-specific semantic frames are first induced by clustering verb uses in a corpus and then verb classes are induced by clustering these frames. By taking this step-wise approach, we can not only generate verb classes based on a massive amount of verb uses in a scalable manner, but also deal with verb polysemy, which is bypassed by most of the previous studies on verb clustering. In our experiments, we acquire semantic frames and verb classes from two giga-word corpora, the larger comprising 20 billion words. The effectiveness of our approach is verified through quantitative evaluations based on polysemy-aware gold-standard data.

Structured Learning for Taxonomy Induction with Belief Propagation
Mohit Bansal, David Burkett, Gerard de Melo, and Dan Klein   2:20–2:45

We present a structured learning approach to inducing hypernym taxonomies using a probabilistic graphical model formulation. Our model incorporates heterogeneous relational evidence about both hypernymy and siblinghood, captured by semantic features based on patterns and statistics from Web n-grams and Wikipedia abstracts. For efficient inference over taxonomy structures, we use loopy belief propagation along with a directed spanning tree algorithm for the core hypernymy factor. To train the system, we extract sub-structures of WordNet and discriminatively learn to reproduce them, using adaptive subgradient stochastic optimization. On the task of reproducing sub-hierarchies of WordNet, our approach achieves a 51% error reduction over a chance baseline, including a 15% error reduction due to the non-hypernym-factored sibling features. On a comparison setup, we find up to 29% relative error reduction over previous work on ancestor F1.

Session 5D: Syntax, Parsing and Tagging IV
Harborside C   *Chair: Phil Blunsom*

A Provably Correct Learning Algorithm for Latent-Variable PCFGs
Shay B. Cohen and Michael Collins   1:30–1:55

We introduce a provably correct learning algorithm for latent-variable PCFGs. The algorithm relies on two steps: first, the use of a matrix-decomposition algorithm applied to a co-occurrence matrix estimated from the parse trees in a training sample; second, the use of EM applied to a convex objective derived from the training samples in combination with the output from the matrix decomposition. Experiments on parsing and a language modeling problem show that the algorithm is efficient and effective in practice.

Spectral Unsupervised Parsing with Additive Tree Metrics
Ankur P. Parikh, Shay B. Cohen, and Eric P. Xing   1:55–2:20

We propose a spectral approach for unsupervised constituent parsing that comes with theoretical guarantees on latent structure recovery. Our approach is grammarless — we directly learn the bracketing structure of a given sentence without using a grammar model. The main algorithm is based on lifting the concept of additive tree metrics for structure learning of latent trees in the phylogenetic and machine learning communities to the case where the tree structure varies across examples. Although finding the "minimal" latent tree is NP-hard in general, for the case of projective trees we find that it can be found using bilexical parsing algorithms. Empirically, our algorithm performs favorably compared to the constituent context model of Klein and Manning (2002) without the need for careful initialization.

[TACL] Cross-lingual Projected Expectation Regularization for Weakly Supervised Learning
Mengqiu Wang and Christopher Manning   2:20–2:45

We consider a multilingual weakly supervised learning scenario where knowledge from annotated corpora in a resource-rich language is transferred via bitext to guide the learning in other languages. Past approaches project labels across bitext and use them as features or gold labels for training. We propose a new method that projects model expectations rather than labels, which facilities transfer of model uncertainty across language boundaries. We encode expectations as constraints and train a discriminative CRF model using Generalized Expectation Criteria (Mann and McCallum, 2010). Evaluated on standard Chinese-English and German-English NER datasets, our method demonstrates F1 scores of 64% and 60% when no labeled data is used. Attaining the same accuracy with supervised CRFs requires 12k and 1.5k labeled sentences. Furthermore, when combined with labeled examples, our method yields significant improvements over state-of-the-art supervised methods, achieving best reported numbers to date on Chinese OntoNotes and German CoNLL-03 datasets.

Session 5E: Cognitive Modeling and Psycholinguistics
Harborside D/E   *Chair: Vera Demberg*

Weak semantic context helps phonetic learning in a model of infant language acquisition
Stella Frank, Naomi H. Feldman, and Sharon Goldwater   1:30–1:55

Learning phonetic categories is one of the first steps to learning a language, yet is hard to do using only distributional phonetic information. Semantics could potentially be useful, since words with different meanings have distinct phonetics, but it is unclear how many word meanings are known to infants learning phonetic categories. We show that attending to a weaker source of semantics, in the form of a distribution over topics in the current context, can lead to improvements in phonetic category learning. In our model, an extension of a previous model of joint word-form and phonetic category inference, the probability of word-forms is topic-dependent, enabling the model to find significantly better phonetic vowel categories and word-forms than a model with no semantic knowledge.

Bootstrapping into Filler-Gap: An Acquisition Story
Marten van Schijndel and Micha Elsner   1:55–2:20

Analyses of filler-gap dependencies usually involve complex syntactic rules or heuristics; however recent results suggest that filler-gap comprehension begins earlier than seemingly simpler constructions such as ditransitives or passives. Therefore, this work models filler-gap acquisition as a byproduct of learning word orderings (e.g. SVO vs OSV), which must be done at a very young age anyway in order to extract meaning from language. Specifically, this model, trained on part-of-speech tags, represents the preferred locations of semantic roles relative to a verb as Gaussian mixtures over real numbers. This approach learns role assignment in filler-gap constructions in a manner consistent with current developmental findings and is extremely robust to initialization variance. Additionally, this model is shown to be able to account for a characteristic error made by learners during this period ('A and B gorped' interpreted as 'A gorped B').

Nonparametric Learning of Phonological Constraints in Optimality Theory
Gabriel Doyle, Klinton Bicknell, and Roger Levy   2:20–2:45

We present a method to jointly learn features and weights directly from distributional data in a maximum entropy model. Specifically, we propose a non-parametric Bayesian model for learning phonological markedness constraints directly from the distribution of input-output mappings in an Optimality Theory (OT) setting. The model uses an Indian Buffet Process prior to learn the feature values used in the maximum entropy method, and is the first algorithm for learning phonological constraints without presupposing constraint structure. The model learns a system of constraints that explains observed data as well as the phonologically-grounded constraints of a standard analysis, with a violation structure corresponding to the standard constraints. These results suggest an alternative data-driven source for constraints instead of a fully innate constraint set.

Parallel Session 6 Overview - Tuesday, June 24, 2014

Track A	Track B	Track C	Track D	Track E	
Machine Translation III	*Lexical Semantics and Ontology II*	*Generation / Summarization / Dialogue*	*NLP Applications and NLP Enabled Technology I*	*Language Resources and Evaluation III*	
Grand V	Grand VI	Harborside A/B	Harborside C	Harborside D/E	
How to Speak a Language without Knowing It *Shi, Knight, and Ji*	Improving sparse word similarity models with asymmetric measures *Gawron*	Single Document Summarization based on Nested Tree Structure *Kikuchi, Hirao, Takamura, Okumura, and Nagata*	Tri-Training for Authorship Attribution with Limited Training Data *Qian, Liu, Chen, and Peng*	Experiments with crowdsourced re-annotation of a POS tagging data set *Hovy, Plank, and Søgaard*	3:15
Assessing the Discourse Factors that Influence the Quality of Machine Translation *Li, Carpuat, and Nenkova*	Dependency-Based Word Embeddings *Levy and Goldberg*	Linguistic Considerations in Automatic Question Generation *Mazidi and Nielsen*	Automation and Evaluation of the Keyword Method for Second Language Learning *Özbal, Pighin, and Strapparava*	Building Sentiment Lexicons for All Major Languages *Chen and Skiena*	3:30
Automatic Detection of Machine Translated Text and Translation Quality Estimation *Aharoni, Koppel, and Goldberg*	Vector spaces for historical linguistics: Using distributional semantics to study syntactic productivity in diachrony *Perek*	Polynomial Time Joint Structural Inference for Sentence Compression *Qian and Liu*	Citation Resolution: A method for evaluating context-based citation recommendation systems *Duma and Klein*	Difficult Cases: From Data to Learning, and Back *Beigman Klebanov and Beigman*	3:45
		A Bayesian Method to Incorporate Background Knowledge during Automatic Text Summarization *Louis*	Hippocratic Abbreviation Expansion *Roark and Sproat*	The VerbCorner Project: Findings from Phase 1 of crowd-sourcing a semantic decomposition of verbs *Hartshorne, Bonial, and Palmer*	4:00
		Predicting Power Relations between Participants in Written Dialog from a Single Thread *Prabhakaran and Rambow*	Unsupervised Feature Learning for Visual Sign Language Identification *Gebre, Crasborn, Wittenburg, Drude, and Heskes*	A Corpus of Sentence-level Revisions in Academic Writing: A Step towards Understanding Statement Strength in Communication *Tan and Lee*	4:15

Parallel Session 6

Session 6A: Machine Translation III
Grand V   *Chair: Niyu Ge*

How to Speak a Language without Knowing It
Xing Shi, Kevin Knight, and Heng Ji   3:15–3:30

We develop a system that lets people overcome language barriers by letting them speak a language they do not know. Our system accepts text entered by a user, translates the text, then converts the translation into a phonetic spelling in the user's own orthography. We trained the system on phonetic spellings in travel phrasebooks.

Assessing the Discourse Factors that Influence the Quality of Machine Translation
Junyi Jessy Li, Marine Carpuat, and Ani Nenkova   3:30–3:45

We present a study of aspects of discourse structure — specifically discourse devices used to organize information in a sentence — that significantly impact the quality of machine translation. Our analysis is based on manual evaluations of translations of news from Chinese and Arabic to English. We find that there is a particularly strong mismatch in the notion of what constitutes a sentence in Chinese and English, which occurs often and is associated with significant degradation in translation quality. Also related to lower translation quality is the need to employ multiple explicit discourse connectives (because, but, etc.), as well as the presence of ambiguous discourse connectives in the English translation. Furthermore, the mismatches between discourse expressions across languages significantly impact translation quality.

Automatic Detection of Machine Translated Text and Translation Quality Estimation
Roee Aharoni, Moshe Koppel, and Yoav Goldberg   3:45–4:00

We show that it is possible to automatically detect machine translated text at sentence level from monolingual corpora, using text classification methods. We show further that the accuracy with which a learned classifier can detect text as machine translated is strongly correlated with the translation quality of the machine translation system that generated it. Finally, we offer a generic machine translation quality estimation technique based on this approach, which does not require reference sentences.

Session 6B: Lexical Semantics and Ontology II
Grand VI   *Chair: Mohit Bansal*

Improving sparse word similarity models with asymmetric measures
Jean Mark Gawron   3:15–3:30

We present several alternative asymmetric models based on Tversky (1977) and show how they improve correlations with human similarity judgments and improve nearest neighbor discovery for both very frequent and very rare words. We attribute this to a principle discovered by Tversky, that in comparing sparse and rich representations, the feature weights may be tipped in favor of recalling the features of the sparser representation, since improvement on our two tasks can be traced to improved performance when comparing mid-frequency and frequent words.

Dependency-Based Word Embeddings
Omer Levy and Yoav Goldberg   3:30–3:45

While continuous word embeddings are gaining popularity, current models are based solely on linear contexts. In this work, we generalize the skip-gram model with negative sampling introduced by Mikolov et al. to include arbitrary contexts. In particular, we perform experiments with dependency-based contexts, and show that they produce markedly different embeddings. The dependency-based embeddings are less topical and exhibit more functional similarity than the original skip-gram embeddings.

Vector spaces for historical linguistics: Using distributional semantics to study syntactic productivity in diachrony
Florent Perek   3:45–4:00

This paper describes an application of distributional semantics to the study of syntactic productivity in diachrony, i.e., the property of grammatical constructions to attract new lexical items over time. By providing an empirical measure of semantic similarity between words derived from lexical co-occurrences, distributional semantics not only reliably captures how the verbs in the distribution of a construction are related, but also enables the use of visualization techniques and statistical modeling to analyze the semantic development of a construction over time and identify the semantic determinants of syntactic productivity in naturally occurring data.

Session 6C: Generation / Summarization / Dialogue
Harborside A/B   *Chair: Asli Celikyilmaz*

Single Document Summarization based on Nested Tree Structure
Yuta Kikuchi, Tsutomu Hirao, Hiroya Takamura, Manabu Okumura, and Masaaki Nagata 3:15–3:30

Many methods of text summarization combining sentence selection and sentence compression have recently been proposed. Although the dependency between words has been used in most of these methods, the dependency between sentences, i.e., rhetorical structures, has not been exploited in such joint methods. We used both dependency between words and dependency between sentences by constructing a nested tree, in which nodes in the document tree representing dependency between sentences were replaced by a sentence tree representing dependency between words. We formulated a summarization task as a combinatorial optimization problem, in which the nested tree was trimmed without losing important content in the source document. The results from an empirical evaluation revealed that our method based on the trimming of the nested tree significantly improved the summarization of texts.

Linguistic Considerations in Automatic Question Generation
Karen Mazidi and Rodney D. Nielsen   3:30–3:45

As students read expository text, comprehension is improved by pausing to answer questions that reinforce the material. We describe an automatic question generator that uses semantic pattern recognition to create questions of varying depth and type for self-study or tutoring. Throughout, we explore how linguistic considerations inform system design. In the described system, semantic role labels of source sentences are used in a domain-independent manner to generate both questions and answers related to the source sentence. Evaluation results show a 44% reduction in the error rate relative to the best prior systems, averaging over all metrics, and up to 61% reduction in the error rate on grammaticality judgments.

Polynomial Time Joint Structural Inference for Sentence Compression
Xian Qian and Yang Liu   3:45–4:00

We propose two polynomial time inference algorithms to compress sentences under bigram and dependency-factored objectives. The first algorithm is exact and requires $O(n^6)$ running time. It extends Eisner's cubic time parsing algorithm by using virtual dependency arcs to link deleted words. Two signatures are added to each span, indicating the number of deleted words and the rightmost kept word within the span. The second algorithm is a fast approximation of the first one. It relaxes the compression ratio constraint using Lagrangian relaxation, and thereby requires $O(n^4)$ running time. Experimental results on the popular sentence compression corpus demonstrate the effectiveness and efficiency of our proposed approach.

A Bayesian Method to Incorporate Background Knowledge during Automatic Text Summarization
Annie Louis   4:00–4:15

In order to summarize a document, it is often useful to have a background set of documents from the domain to serve as a reference for determining new and important information in the input document. We present a model based on Bayesian surprise which provides an intuitive way to identify surprising information from summarization input with respect to a background corpus. Specifically, the method quantifies the degree to which pieces of information in the input change one's beliefs' about the world represented in the background. We develop systems for generic and update summarization based on this idea. Our method provides competitive content selection performance with particular advantages in the update task where systems are given a small and topical background corpus.

Predicting Power Relations between Participants in Written Dialog from a Single Thread
Vinodkumar Prabhakaran and Owen Rambow   4:15–4:30

We introduce the problem of predicting who has power over whom in pairs of people based on a single written dialog. We propose a new set of structural features. We build a supervised learning system to predict the direction of power; our new features significantly improve the results over using previously proposed features.

Session 6D: NLP Applications and NLP Enabled Technology I
Harborside C   *Chair: Hwee Tou Ng*

Tri-Training for Authorship Attribution with Limited Training Data
Tieyun Qian, Bing Liu, Li Chen, and Zhiyong Peng   3:15–3:30

Authorship attribution (AA) aims to identify the authors of a set of documents. Traditional studies in this area often assume that there are a large set of labeled documents available for training. However, in the real life, it is often difficult or expensive to collect a large set of labeled data. For example, in the online review domain, most reviewers (authors) only write a few reviews, which are not enough to serve as the training data for accurate classification. In this paper, we present a novel three-view tri-training method to iteratively identify authors of unlabeled data to augment the training set. The key idea is to first represent each docu-ment in three distinct views, and then perform tri-training to exploit the large amount of unlabeled documents. Starting from 10 train-ing documents per author, we systematically evaluate the effectiveness of the proposed tri-training method for AA. Experimental results show that the proposed approach outperforms the state-of-the-art semi-supervised method CNG+SVM and other baselines.

Automation and Evaluation of the Keyword Method for Second Language Learning
Gözde Özbal, Daniele Pighin, and Carlo Strapparava   3:30–3:45

In this paper, we combine existing NLP techniques with minimal supervision to build memory tips according to the keyword method, a well established mnemonic device for second language learning. We present what we believe to be the first extrinsic evaluation of a creative sentence generator on a vocabulary learning task. The results demonstrate that NLP techniques can effectively support the development of resources for second language learning.

Citation Resolution: A method for evaluating context-based citation recommendation systems
Daniel Duma and Ewan Klein   3:45–4:00

Wouldn't it be helpful if your text editor automatically suggested papers that are relevant to your research? Wouldn't it be even better if those suggestions were contextually relevant? In this paper we name a system that would accomplish this a context-based citation recommendation (CBCR) system. We specifically present Citation Resolution, a method for the evaluation of CBCR systems which exclusively uses readily-available scientific articles. Exploiting the human judgements that are already implicit in available resources, we avoid purpose-specific annotation. We apply this evaluation to three sets of methods for representing a document, based on a) the contents of the document, b) the surrounding contexts of citations to the document found in other documents, and c) a mixture of the two.

Hippocratic Abbreviation Expansion
Brian Roark and Richard Sproat   4:00–4:15

Incorrect normalization of text can be particularly damaging for applications like text-to-speech synthesis (TTS) or typing auto-correction, where the resulting normalization is directly presented to the user, versus feeding downstream applications. In this paper, we focus on abbreviation expansion for TTS, which requires a "do no harm", high precision approach yielding few expansion errors at the cost of leaving relatively many abbreviations unexpanded. In the context of a large-scale, real-world TTS scenario, we present methods for training classifiers to establish whether a particular expansion is apt. We achieve a large increase in correct abbreviation expansion when combined with the baseline text normalization component of the TTS system, together with a substantial reduction in incorrect expansions.

Unsupervised Feature Learning for Visual Sign Language Identification
Binyam Gebrekidan Gebre, Onno Crasborn, Peter Wittenburg, Sebastian Drude, and Tom Heskes
4:15–4:30

Prior research on language identification focused primarily on text and speech. In this paper, we focus on the visual modality and present a method for identifying sign languages solely from short video samples. The method is trained on unlabelled video data (unsupervised feature learning) and using these features, it is trained to discriminate between six sign languages (supervised learning). We ran experiments on short video samples involving 30 signers (about 6 hours in total). Using leave-one-signer-out cross-validation, our evaluation shows an average best accuracy of 84%. Given that sign languages are under-resourced, unsupervised feature learning techniques are the right tools and our results indicate that this is realistic for sign language identification.

Session 6E: Language Resources and Evaluation III
Harborside D/E   *Chair: Roberto Navigli*

Experiments with crowdsourced re-annotation of a POS tagging data set
Dirk Hovy, Barbara Plank, and Anders Søgaard   3:15–3:30

Crowdsourcing lets us collect multiple annotations for an item from several annotators. Typically, these are annotations for non-sequential classification tasks. While there has been some work on crowdsourcing named entity annotations, researchers have largely assumed that syntactic tasks such as part-of-speech (POS) tagging cannot be crowdsourced. This paper shows that workers can actually annotate sequential data almost as well as experts. Fur- ther, we show that the models learned from crowdsourced annotations fare as well as the models learned from expert annotations in downstream tasks.

Building Sentiment Lexicons for All Major Languages
Yanqing Chen and Steven Skiena   3:30–3:45

Sentiment analysis in a multilingual world remains a challenging problem, because developing language-specific sentiment lexicons is an extremely resource-intensive process. Such lexicons remain a scarce resource for most languages. In this paper, we address this lexicon gap by building high-quality sentiment lexicons for 136 major languages. We integrate a variety of linguistic resources to produce an immense knowledge graph. By appropriately propagating from seed words, we construct sentiment lexicons for each component language of our graph. Our lexicons have a polarity agreement of 95.7% with published lexicons, while achieving an overall coverage of 45.2%. We demonstrate the performance of our lexicons in an extrinsic analysis of 2,000 distinct historical figures' Wikipedia articles on 30 languages. Despite cultural difference and the intended neutrality of Wikipedia articles, our lexicons show an average sentiment correlation of 0.28 across all language pairs.

Difficult Cases: From Data to Learning, and Back
Beata Beigman Klebanov and Eyal Beigman   3:45–4:00

This article contributes to the ongoing discussion in the computational linguistics community regarding instances that are difficult to annotate reliably. Is it worthwhile to identify those? What information can be inferred from them regarding the nature of the task? What should be done with them when building supervised machine learning systems? We address these questions in the context of a subjective semantic task. In this setting, we show that the presence of such instances in training data misleads a machine learner into misclassifying clear-cut cases. We also show that considering machine learning outcomes with and without the difficult cases, it is possible to identify specific weaknesses of the problem representation.

The VerbCorner Project: Findings from Phase 1 of crowd-sourcing a semantic decomposition of verbs
Joshua K. Hartshorne, Claire Bonial, and Martha Palmer   4:00–4:15

Any given verb can appear in some syntactic frames (Sally broke the vase, The vase broke) but not others (*Sally broke at the vase, Sally broke the vase to John). There is now considerable evidence that the syntactic behaviors of some verbs can predicted by their meanings, and many current theories posit that this is true for most if not all verbs. If true, this fact would have striking implications for theories and models of language acquisition, as well as numerous applications in natural language processing. However, empirical investigations to date have focused on a small number of verbs. We report on early results from VerbCorner, a crowd-sourced project extending this work to a large, representative sample of English verbs.

A Corpus of Sentence-level Revisions in Academic Writing: A Step towards Understanding Statement Strength in Communication
Chenhao Tan and Lillian Lee   4:15–4:30

The strength with which a statement is made can have a significant impact on the audience. For example, international relations can be strained by how the media in one country describes an event in another; and papers can be rejected because they overstate or understate their findings. It is thus important to understand the effects of statement strength. A first step is to be able to distinguish between strong and weak statements. However, even this problem is understudied, partly due to a lack of data. Since strength is inherently relative, revisions of texts that make claims are a natural source of data on strength differences. In this paper, we introduce a corpus of sentence-level revisions from academic writing. We also describe insights gained from our annotation efforts for this task.

Posters: Long papers

Active Learning with Efficient Feature Weighting Methods for Improving Data Quality and Classification Accuracy
Justin Martineau, Lu Chen, Doreen Cheng, and Amit Sheth

Many machine learning datasets are noisy with a substantial number of mislabeled instances. This noise yields sub-optimal classification performance. In this paper we study a large, low quality annotated dataset, created quickly and cheaply using Amazon Mechanical Turk to crowdsource annotations. We describe computationally cheap feature weighting techniques and a novel non-linear distribution spreading algorithm that can be used to iteratively and interactively correcting mislabeled instances to significantly improve annotation quality at low cost. Eight different emotion extraction experiments on Twitter data demonstrate that our approach is just as effective as more computationally expensive techniques. Our techniques save a considerable amount of time.

Political Ideology Detection Using Recursive Neural Networks
Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik

An individual's words often reveal their political ideology. Existing automated techniques to identify ideology from text focus on bags of words or wordlists, ignoring syntax. Taking inspiration from recent work in sentiment analysis that successfully models the compositional aspect of language, we apply a recursive neural network (RNN) framework to the task of identifying the political position evinced by a sentence. To demonstrate the effectiveness of this method with additional subsentential elements, we crowdsource political annotations at a phrase and sentence level. We demonstrate that our model outperforms existing models on our new dataset and an existing dataset.

A Unified Model for Soft Linguistic Reordering Constraints in Statistical Machine Translation
Junhui Li, Yuval Marton, Philip Resnik, and Hal Daumé III

This paper explores a simple and effective unified framework for incorporating soft linguistic reordering constraints into a hierarchical phrase-based translation system: 1) a syntactic reordering model that explores reorderings for context free grammar rules; and 2) a semantic reordering model that focuses on the reordering of predicate-argument structures. We develop novel features based on both models and use them as soft constraints to guide the translation process. Experiments on Chinese-English translation show that the reordering approach can significantly improve a state-of-the-art hierarchical phrase-based translation system. However, the gain achieved by the semantic reordering model is limited in the presence of the syntactic reordering model, and we therefore provide a detailed analysis of the behavior differences between the two.

Are Two Heads Better than One? Crowdsourced Translation via a Two-Step Collaboration of Non-Professional Translators and Editors
Rui Yan, Mingkun Gao, Ellie Pavlick, and Chris Callison-Burch

Crowdsourcing is a viable mechanism for creating training data for machine translation. It provides a low cost, fast turnaround way of processing large volumes of data. However, when compared to professional translation, naive collection of translations from non-professionals yields low-quality results. Careful quality control is necessary for crowdsourcing to work well. In this paper, we examine the challenges of a two-step collaboration process with translation and post-editing by non-professionals. We develop graph-based ranking models that automatically select the best output from multiple redundant versions of translations and edits, and improves translation quality closer to professionals.

A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing
Rene Pickhardt, Thomas Gottron, Martin Körner, Paul Georg Wagner, Till Speicher, and Steffen Staab

We introduce a novel approach for building language models based on a systematic, recursive exploration of skip n-gram models which are interpolated using modified Kneser Ney smoothing. Our approach generalizes language models as it contains the classical interpolation with lower order models as a special case. In this paper we motivate, formalize and present our approach. In an extensive empirical experiment we demonstrate that our generalized language models lead to a substantial reduction of perplexity between 3.1% and 12.7% in comparison to traditional language models using modified Kneser Ney smoothing. Furthermore, we inves-

tigate the behaviour over three other languages and a domain specific corpus where we observed consistent improvements. Finally, we also show that the strength of our approach lies in its ability to cope in particular with sparse training data. Using a very small training data set of only 736 KB text we yield improvements of even 25.7% reduction of perplexity.

A Semiparametric Gaussian Copula Regression Model for Predicting Financial Risks from Earnings Calls
William Yang Wang and Zhenhao Hua

Earnings call summarizes the financial performance of a company, and it is an important indicator of the future financial risks of the company. We quantitatively study how earnings calls are correlated with the financial risks, with a special focus on the financial crisis of 2009. In particular, we perform a text regression task: given the transcript of an earnings call, we predict the volatility of stock prices from the week after the call is made. We propose the use of copula: a powerful statistical framework that separately models the uniform marginals and their complex multivariate stochastic dependencies, while not requiring any prior assumptions on the distributions of the covariate and the dependent variable. By performing probability integral transform, our approach moves beyond the standard count-based bag-of-words models in NLP, and improves previous work on text regression by incorporating the correlation among local features in the form of semiparametric Gaussian copula. In experiments, we show that our model significantly outperforms strong linear and non-linear discriminative baselines on three datasets under various settings.

Polylingual Tree-Based Topic Models for Translation Domain Adaptation
Yuening Hu, Ke Zhai, Vladimir Eidelman, and Jordan Boyd-Graber

Topic models, an unsupervised technique for inferring translation domains improve machine translation quality. However, previous work uses only the source language and completely ignores the target language, which can disambiguate domains. We propose new polylingual tree-based topic models to extract domain knowledge that considers both source and target languages and derive three different inference schemes. We evaluate our model on a Chinese to English translation task and obtain up to 1.2 BLEU improvement over strong baselines.

Low-Resource Semantic Role Labeling
Matthew R. Gormley, Margaret Mitchell, Benjamin Van Durme, and Mark Dredze

We explore the extent to which high-resource manual annotations such as treebanks are necessary for the task of semantic role labeling (SRL). We examine how performance changes without syntactic supervision, comparing both joint and pipelined methods to induce latent syntax. This work highlights a new application of unsupervised grammar induction and demonstrates several approaches to SRL in the absence of supervised syntax. Our best models obtain competitive results in the high-resource setting and state-of-the-art results in the low resource setting, reaching 72.48% F1 averaged across languages. We release our code for this work along with a larger toolkit for specifying arbitrary graphical structure.

Joint Syntactic and Semantic Parsing with Combinatory Categorial Grammar
Jayant Krishnamurthy and Tom M. Mitchell

We present an approach to training a joint syntactic and semantic parser that combines syntactic training information from CCGbank with semantic training information from a knowledge base via distant supervision. The trained parser produces a full syntactic parse of any sentence, while simultaneously producing logical forms for portions of the sentence that have a semantic representation within the parser's predicate vocabulary. We demonstrate our approach by training a parser whose semantic representation contains 130 predicates from the NELL ontology. A semantic evaluation demonstrates that this parser produces logical forms better than both comparable prior work and a pipelined syntax-then-semantics approach. A syntactic evaluation on CCGbank demonstrates that the parser's dependency F- score is within 2.5% of state-of-the-art.

Learning Semantic Hierarchies via Word Embeddings
Ruiji Fu, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu

Semantic hierarchy construction aims to build structures of concepts linked by hypernym—hyponym ("is-a") relations. A major challenge for this task is the automatic discovery of such relations. This paper proposes a novel and effective method for the construction of semantic hierarchies based on word embeddings, which can be used to measure the semantic relationship between words. We identify whether a candidate word pair has hypernym—hyponym relation by using the word-embedding-based semantic projections between words and their hypernyms. Our result, an F-score of 73.74%, outperforms the state-of-theart methods on a manually

labeled test dataset. Moreover, combining our method with a previous manually-built hierachy extension method can further improve Fscore to 80.29%.

Probabilistic Soft Logic for Semantic Textual Similarity
Islam Beltagy, Katrin Erk, and Raymond Mooney

Probabilistic Soft Logic (PSL) is a recently developed framework for probabilistic logic. We use PSL to combine logical and distributional representations of natural-language meaning, where distributional information is represented in the form of weighted inference rules. We apply this framework to the task of Semantic Textual Similarity (STS) (i.e. judging the semantic similarity of natural language sentences), and show that PSL gives improved results compared to a previous approach based on Markov Logic Networks (MLNs) and a purely distributional approach.

Abstractive Summarization of Spoken and Written Conversations Based on Phrasal Queries
Yashar Mehdad, Giuseppe Carenini, and Raymond T. Ng

We propose a novel abstractive query-based summarization system for conversations, where queries are defined as phrases reflecting a user information needs. We rank and extract the utterances in a conversation based on the overall content and the phrasal query information. We cluster the selected sentences based on their lexical similarity and aggregate the sentences in each cluster by means of a word graph model. We propose a ranking strategy to select the best path in the constructed graph as a query-based abstract sentence for each cluster. A resulting summary consists of abstractive sentences representing the phrasal query information and the overall content of the conversation. Automatic and manual evaluation results over meeting, chat and email conversations show that our approach significantly outperforms baselines and previous extractive models.

Comparing Multi-label Classification with Reinforcement Learning for Summarisation of Time-series Data
Dimitra Gkatzia, Helen Hastie, and Oliver Lemon

We present a novel approach for automatic report generation from time-series data, in the context of student feedback generation. Our proposed methodology treats content selection as a multi-label (ML) classification problem, which takes as input time-series data and outputs a set of templates, while capturing the dependencies between selected templates. We show that this method generates output closer to the feedback that lecturers actually generated, achieving 3.5% higher accuracy and 15% higher F-score than multiple simple classifiers that keep a history of selected templates. Furthermore, we compare a ML classifier with a Reinforcement Learning (RL) approach in simulation and using ratings from real student users. We show that the different methods have different benefits, with ML being more accurate for predicting what was seen in the training data, whereas RL is more exploratory and slightly preferred by the students.

Approximation Strategies for Multi-Structure Sentence Compression
Kapil Thadani

Sentence compression has been shown to benefit from joint inference over n-gram and dependency-based objectives but this typically requires expensive integer programming. We explore instead the use of Lagrangian relaxation to decouple the two subproblems and solve them separately. While dynamic programming is viable for bigram-based sentence compression, finding optimal compressed trees in graphs is NP-hard. We recover approximate solutions to this problem using LP relaxation and maximum spanning tree algorithms, yielding techniques that can be combined with the efficient bigram-based inference approach using Lagrange multipliers. Experiments show that these approximation strategies produce results comparable to a state-of-the-art integer linear programming formulation for the same joint inference task along with a significant improvement in runtime.

Opinion Mining on YouTube
Aliaksei Severyn, Alessandro Moschitti, Olga Uryupina, Barbara Plank, and Katja Filippova

This paper defines a systematic approach to Opinion Mining (OM) on YouTube comments by (i) modeling classifiers for predicting the opinion polarity and the type of comment and (ii) proposing robust shallow syntactic structures for improving model adaptability. We rely on the tree kernel technology to automatically extract and learn features with better generalization power than bag-of-words. An extensive empirical evaluation on our manually annotated YouTube comments corpus shows a high classification accuracy and highlights the benefits of structural models in a cross-domain setting.

Automatic Keyphrase Extraction: A Survey of the State of the Art
Kazi Saidul Hasan and Vincent Ng

We present a survey of the state of the art in keyphrase extraction.

Pattern Dictionary of English Prepositions
Ken Litkowski

We present a new lexical resource for the study of preposition behavior, the Pattern Dictionary of English Prepositions (PDEP). This dictionary, which follows principles laid out in Hanks' theory of norms and exploitations, is linked to 81,509 sentences for 304 prepositions, which have been made available under The Preposition Project (TPP). Notably, 47,285 sentences, initially untagged, provide a representative sample of preposition use, unlike the tagged sentences used in previous studies. Each sentence has been parsed with a dependency parser and our system has near-instantaneous access to features developed with this parser to explore and annotate properties of individual senses. The features make extensive use of WordNet. We have extended feature exploration to include lookup of FrameNet lexical units and VerbNet classes for use in characterizing preposition behavior. We have designed our system to allow public access to any of the data available in the system.

Looking at Unbalanced Specialized Comparable Corpora for Bilingual Lexicon Extraction
Emmanuel Morin and Amir Hazem

The main work in bilingual lexicon extraction from comparable corpora is based on the implicit hypothesis that corpora are balanced. However, the historical context-based projection method dedicated to this task is relatively insensitive to the sizes of each part of the comparable corpus. Within this context, we have carried out a study on the influence of unbalanced specialized comparable corpora on the quality of bilingual terminology extraction through different experiments. Moreover, we have introduced a regression model that boosts the observations of word co-occurrences used in the context-based projection method. Our results show that the use of unbalanced specialized comparable corpora induces a significant gain in the quality of extracted lexicons.

Validating and Extending Semantic Knowledge Bases using Video Games with a Purpose
Daniele Vannella, David Jurgens, Daniele Scarfini, Domenico Toscani, and Roberto Navigli

Large-scale knowledge bases are important assets in NLP. Frequently, such resources are constructed through automatic mergers of complementary resources, such as WordNet and Wikipedia. However, manually validating these resources is prohibitively expensive, even when using methods such as crowdsourcing. We propose a cost-effective method of validating and extending knowledge bases using video games with a purpose. Two video games were created to validate concept-concept and concept-image relations. In experiments comparing with crowdsourcing, we show that video game-based validation consistently leads to higher-quality annotations, even when players are not compensated.

Shallow Analysis Based Assessment of Syntactic Complexity for Automated Speech Scoring
Suma Bhat, Huichao Xue, and Su-Youn Yoon

Designing measures that capture various aspects of language ability is a central task in the design of systems for automatic scoring of spontaneous speech. In this study, we address a key aspect of language proficiency assessment – syntactic complexity. We propose a novel measure of syntactic complexity for spontaneous speech that shows optimum empirical performance on real world data in multiple ways. First, it is both *robust* and *reliable*, producing automatic scores that agree well with human rating compared to the state-of-the-art. Second, the measure makes sense theoretically, both from algorithmic and native language acquisition points of view.

Can You Repeat That? Using Word Repetition to Improve Spoken Term Detection
Jonathan Wintrode and Sanjeev Khudanpur

We aim to improve spoken term detection performance by incorporating contextual information beyond traditional N-gram language models. Instead of taking a broad view of topic context in spoken documents, variability of word co-occurrence statistics across corpora leads us to focus instead the on phenomenon of word repetition within single documents. We show that given the detection of one instance of a term we are more likely to detect additional instances of that term in the same document. We leverage term 'burstiness' for each keyword by taking the most confident hypothesis in each document and interpolating with lower scoring hits. We then develop a principled approach to select interpolation weights using only the ASR training data. Using our re-weighting approach we demonstrate up to 1.5% absolute performance improvement on the term

detection task in all Babel program languages.

Character-Level Chinese Dependency Parsing
Meishan Zhang, Yue Zhang, Wanxiang Che, and Ting Liu

Recent work on Chinese analysis has led to large-scale annotations of the internal structures of words, enabling character-level analysis of Chinese syntactic structures. In this paper, we investigate the problem of character-level Chinese dependency parsing, building dependency trees over characters. Character-level information can benefit downstream applications by offering flexible granularities for word segmentation while improving word-level dependency parsing accuracies. We present novel adaptations of two major shift-reduce dependency parsing algorithms to character-level parsing. Experimental results on the Chinese Treebank demonstrate improved performances over word-based parsing methods.

Unsupervised Dependency Parsing with Transferring Distribution via Parallel Guidance and Entropy Regularization
Xuezhe Ma and Fei Xia

We present a novel approach for inducing unsupervised dependency parsers for languages that have no labeled training data, but have translated text in a resource-rich language. We train probabilistic parsing models for resource-poor languages by transferring cross-lingual knowledge from resource-rich language with entropy regularization. Our method can be used as a purely monolingual dependency parser, requiring no human translations for the test data, thus making it applicable to a wide range of resource-poor languages. We perform experiments on three Data sets — Version 1.0 and version 2.0 of Google Universal Dependency Treebanks and Treebanks from CoNLL shared-tasks, across ten languages. We obtain state-of-the art performance of all the three data sets when compared with previously studied unsupervised and projected parsing systems.

Unsupervised Morphology-Based Vocabulary Expansion
Mohammad Sadegh Rasooli, Thomas Lippincott, Nizar Habash, and Owen Rambow

We present a novel way of generating unseen words, which is useful for certain applications such as automatic speech recognition or optical character recognition in low-resource languages. We test our vocabulary generator on seven low-resource languages by measuring the decrease in out-of-vocabulary word rate on a held-out test set. The languages we study have very different morphological properties; we show how our results differ depending on the morphological complexity of the language. In our best result (on Assamese), our approach can predict 29% of the token-based out-of-vocabulary with a small amount of unlabeled training data.

Toward Better Chinese Word Segmentation for SMT via Bilingual Constraints
Xiaodong Zeng, Lidia S. Chao, Derek F. Wong, Isabel Trancoso, and Liang Tian

This study investigates on building a better Chinese word segmentation model for statistical machine translation. It aims at leveraging word boundary information, automatically learned by bilingual character-based alignments, to induce a preferable segmentation model. We propose dealing with the induced word boundaries as soft constraints to bias the continuous learning of a supervised CRFs model, trained by the treebank data (labeled), on the bilingual data (unlabeled). The induced word boundary information is encoded as a graph propagation constraint. The constrained model induction is accomplished by using posterior regularization algorithm. The experiments on a Chinese-to-English machine translation task reveal that the proposed model can bring positive segmentation effects to translation quality.

Posters: Short papers

Determiner-Established Deixis to Communicative Artifacts in Pedagogical Text
Shomir Wilson and Jon Oberlander

Pedagogical materials frequently contain deixis to communicative artifacts such as textual structures (e.g., sections and lists), discourse entities, and illustrations. By relating such artifacts to the prose, deixis plays an essential role in structuring the flow of information in informative writing. However, existing language technologies have largely overlooked this mechanism. We examine properties of deixis to communicative artifacts using a corpus rich in determiner-established instances of the phenomenon (e.g., "this section", "these equations", "those reasons") from Wikibooks, a collection of learning texts. We use this corpus in combination with WordNet to determine a set of word senses that are characteristic of the phenomenon, showing its diversity

and validating intuitions about its qualities. The results motivate further research to extract the connections encoded by such deixis, with the goals of enhancing tools to present pedagogical e-texts to readers and, more broadly, improving language technologies that rely on deictic phenomena.

Modeling Factuality Judgments in Social Media Text
Sandeep Soni, Tanushree Mitra, Eric Gilbert, and Jacob Eisenstein

How do journalists mark quoted content as certain or uncertain, and how do readers interpret these signals? Predicates such as "thinks", "claims", and "admits" offer a range of options for framing quoted content according to the author's own perceptions of its credibility. We gather a new dataset of direct and indirect quotes from Twitter, and obtain annotations of the perceived certainty of the quoted statements. We then compare the ability of linguistic and extra-linguistic features to predict readers' assessment of the certainty of quoted content. We see that readers are indeed influenced by such framing devices — and we find no evidence that they consider other factors, such as the source, journalist, or the content itself. In addition, we examine the impact of specific framing devices on perceptions of credibility.

A Topic Model for Building Fine-grained Domain-specific Emotion Lexicon
Min Yang, Dingju Zhu, and Kam-Pui Chow

Emotion lexicons play a crucial role in sentiment analysis and opinion mining. In this paper, we propose a novel Emotion-aware LDA (EaLDA) model to build a domain-specific lexicon for predefined emotions that include anger, disgust, fear, joy, sadness, surprise. The model uses a minimal set of domain-independent seed words as prior knowledge to discover a domain-specific lexicon, learning a fine-grained emotion lexicon much richer and adaptive to a specific domain. By comprehensive experiments, we show that our model can generate a high-quality fine-grained domain-specific emotion lexicon.

Depeche Mood: a Lexicon for Emotion Analysis from Crowd Annotated News
Jacopo Staiano and Marco Guerini

While many lexica annotated with words polarity are available for sentiment analysis, very few tackle the harder task of emotion analysis and are usually quite limited in coverage. In this paper, we present a novel approach for extracting - in a totally automated way - a high-coverage and high-precision lexicon of roughly 37 thousand terms annotated with emotion scores, called DepecheMood. Our approach exploits in an original way 'crowd-sourced' affective annotation implicitly provided by readers of news articles from rappler.com. By providing new state-of-the-art performances in unsupervised settings for regression and classification tasks, even using a naive approach, our experiments show the beneficial impact of harvesting social media data for affective lexicon building

Improving Twitter Sentiment Analysis with Topic-Based Mixture Modeling and Semi-Supervised Training
Bing Xiang and Liang Zhou

In this paper, we present multiple approaches to improve sentiment analysis on Twitter data. We first establish a state-of-the-art baseline with a rich feature set. Then we build a topic-based sentiment mixture model with topic-specific data in a semi-supervised training framework. The topic information is generated through topic modeling based on an efficient implementation of Latent Dirichlet Allocation (LDA). The proposed sentiment model outperforms the top system in the task of Sentiment Analysis in Twitter in SemEval-2013 in terms of averaged F scores.

Cross-cultural Deception Detection
Verónica Pérez-Rosas and Rada Mihalcea

In this paper, we address the task of cross-cultural deception detection. Using crowdsourcing, we collect three deception datasets, two in English (one originating from United States and one from India), and one in Spanish obtained from speakers from Mexico. We run comparative experiments to evaluate the accuracies of deception classifiers built for each culture, and also to analyze classification differences within and across cultures. Our results show that we can leverage cross-cultural information, either through translation or equivalent semantic categories, and build deception classifiers with a performance ranging between 60-70%.

Particle Filter Rejuvenation and Latent Dirichlet Allocation
Chandler May, Alex Clemmer, and Benjamin Van Durme

Previous research has established several methods of online learning for latent Dirichlet allocation (LDA). However, streaming learning for LDA—allowing only one pass over the data and constant storage complexity—

is not as well explored. We use reservoir sampling to reduce the storage complexity of a previously-studied online algorithm, namely the particle filter, to constant. We then show that a simpler particle filter implementation performs just as well, and that the quality of the initialization dominates other factors of performance.

Comparing Automatic Evaluation Measures for Image Description
Desmond Elliott and Frank Keller

Image description is a new natural language generation task, where the aim is to generate a human-like description of an image. The evaluation of computer-generated text is a notoriously difficult problem, however, the quality of image descriptions has typically been measured using unigram BLEU and human judgements. The focus of this paper is to determine the correlation of automatic measures with human judgements for this task. We estimate the correlation of unigram and Smoothed BLEU, TER, ROUGE-SU4, and Meteor against human judgements on two data sets. The main finding is that unigram BLEU has a weak correlation, and Meteor has the strongest correlation with human judgements.

Learning a Lexical Simplifier Using Wikipedia
Colby Horn, Cathryn Manduca, and David Kauchak

In this paper we introduce a new lexical simplification approach. We extract over 30K candidate lexical simplifications by identifying aligned words in a sentence-aligned corpus of English Wikipedia with Simple English Wikipedia. To apply these rules, we learn a feature-based ranker using SVMrank trained on a set of labeled simplifications collected using Amazon's Mechanical Turk. Using human simplifications for evaluation, we achieve a precision of 76% with changes in 86% of the examples.

Cheap and easy entity evaluation
Ben Hachey, Joel Nothman, and Will Radford

The AIDA-YAGO dataset is a popular target for whole-document entity recognition and disambiguation, despite lacking a shared evaluation tool. We review evaluation regimens in the literature while comparing the output of three approaches, and identify research opportunities. This utilises our open, accessible evaluation tool. We exemplify a new paradigm of distributed, shared evaluation, in which evaluation software and standardised, versioned system outputs are provided online.

Identifying Real-Life Complex Task Names with Task-Intrinsic Entities from Microblogs
Ting-Xuan Wang, Kun-Yu Tsai, and Wen-Hsiang Lu

Recently, users who search on the web are targeting to more complex tasks due to the explosive growth of web usage. To accomplish a complex task, users may need to obtain information of various entities. For example, a user who wants to plan a trip to Beijing, should book a flight, reserve a hotel room, and survey a Beijing map. A complex task thus needs to submit several queries in order to seeking each of entities. Understanding complex tasks can allow a search engine to suggest related entities and help users explicitly assign their ongoing tasks.

Mutual Disambiguation for Entity Linking
Eric Charton, Marie-Jean Meurs, Ludovic Jean-Louis, and Michel Gagnon

The disambiguation algorithm presented in this paper is implemented in SemLinker, an entity linking system. First, named entities are linked to candidate Wikipedia pages by a generic annotation engine. Then, the algorithm re-ranks candidate links according to mutual relations between all the named entities found in the document. The evaluation is based on experiments conducted on the test corpus of the TAC-KBP 2012 entity linking task.

How Well can We Learn Interpretable Entity Types from Text?
Dirk Hovy

Many NLP applications rely on type systems to represent higher-level classes. Domain-specific ones are more informative, but have to be manually tailored to each task and domain, making them inflexible and expensive. We investigate a largely unsupervised approach to learning interpretable, domain-specific entity types from unlabeled text. It assumes that any common noun in a domain can function as potential entity type, and uses those nouns as hidden variables in a HMM. To constrain training, it extracts co-occurrence dictionaries of entities and common nouns from the data. We evaluate the learned types by measuring their prediction accuracy for verb arguments in several domains. The results suggest that it is possible to learn domain-specific entity types from unlabeled data. We show significant improvements over an informed baseline, reducing the error rate by 56%.

Learning Translational and Knowledge-based Similarities from Relevance Rankings for Cross-Language Retrieval
Shigehiko Schamoni, Felix Hieber, Artem Sokolov, and Stefan Riezler

We present an approach to cross-language retrieval that combines dense knowledge-based features and sparse word translations. Both feature types are learned directly from relevance rankings of bilingual documents in a pairwise ranking framework. In large-scale experiments for patent prior art search and cross-lingual retrieval in Wikipedia, our approach yields considerable improvements over learning-to-rank with either only dense or only sparse features, and over very competitive baselines that combine state-of-the-art machine translation and retrieval.

Two-Stage Hashing for Fast Document Retrieval
Hao Li, Wei Liu, and Heng Ji

This work fulfills sublinear time Nearest Neighbor Search (NNS) in massive scale document collections. The primary contribution is to propose a two-stage unsupervised hashing framework which harmoniously integrates two state-of-the-art hashing algorithms Locality Sensitive Hashing (LSH) and Iterative Quantization (ITQ). LSH accounts for neighbor candidate pruning, while ITQ provides an efficient and effective reranking over the neighbor pool captured by LSH. Furthermore, the proposed hashing framework capitalizes on both term and topic similarity among documents, leading to precise document retrieval. The experimental results convincingly show that our hashing based document retrieval approach well approximates the conventional Information Retrieval (IR) method in terms of retrieving semantically similar documents, and meanwhile achieves a speedup of over one order of magnitude in query time.

An Annotation Framework for Dense Event Ordering
Taylor Cassidy, Bill McDowell, Nathanael Chambers, and Steven Bethard

Today's event ordering research is heavily dependent on annotated corpora. Current corpora influence shared evaluations and drive algorithm development. Partly due to this dependence, most research focuses on partial orderings of a document's events. For instance, the TempEval competitions and the TimeBank only annotate small portions of the event graph, focusing on the most salient events or on specific types of event pairs (e.g., only events in the same sentence). Deeper temporal reasoners struggle with this sparsity because the entire temporal picture is not represented. This paper proposes a new annotation process with a mechanism to force annotators to label connected graphs. It generates 10 times more relations per document than the TimeBank, and our TimeBank-Dense corpus is larger than all current corpora. We hope this process and its dense corpus encourages research on new global models with deeper reasoning.

Linguistically debatable or just plain wrong?
Barbara Plank, Dirk Hovy, and Anders Søgaard

In linguistic annotation projects, we typically develop annotation guidelines to minimize disagreement. However, in this position paper we question whether we should actually limit the disagreements between annotators, rather than embracing them. We present an empirical analysis of part-of-speech annotated data sets that suggests that disagreements are systematic across domains and to a certain extend also across languages. This points to an underlying ambiguity rather than random errors. Moreover, a quantitative analysis of tag confusions reveals that the majority of disagreements are due to linguistically debatable cases rather than annotation errors. Specifically, we show that even in the absence of annotation guidelines only 2% of annotator choices are linguistically unmotivated.

Humans Require Context to Infer Ironic Intent (so Computers Probably do, too)
Byron C. Wallace, Do Kook Choe, Laura Kertz, and Eugene Charniak

Automatically detecting verbal irony (roughly, sarcasm) is a challenging task because ironists say something other than — and often opposite to — what they actually mean. Discerning ironic intent exclusively from the words and syntax comprising texts (e.g., tweets, forum posts) is therefore not always possible: additional contextual information about the speaker and/or the topic at hand is often necessary. We introduce a new corpus that provides empirical evidence for this claim. We show that annotators frequently require context to make judgements concerning ironic intent, and that machine learning approaches tend to misclassify those same comments for which annotators required additional context.

Automatic prediction of aspectual class of verbs in context
Annemarie Friedrich and Alexis Palmer

This paper describes a new approach to predicting the aspectual class of verbs in context, i.e., whether a verb is used in a stative or dynamic sense. We identify two challenging cases of this problem: when the verb is unseen in training data, and when the verb is ambiguous for aspectual class. A semi-supervised approach using linguistically-motivated features and a novel set of distributional features based on representative verb types allows us to predict classes accurately, even for unseen verbs. Many frequent verbs can be either stative or dynamic in different contexts, which has not been modeled by previous work; we use contextual features to resolve this ambiguity. In addition, we introduce two new datasets of clauses marked for aspectual class.

Combining Word Patterns and Discourse Markers for Paradigmatic Relation Classification
Michael Roth and Sabine Schulte im Walde

Distinguishing between paradigmatic relations such as synonymy, antonymy and hypernymy is an important prerequisite in a range of NLP applications. In this paper, we explore discourse relations as an alternative set of features to lexico-syntactic patterns. We demonstrate that statistics over discourse relations, collected via explicit discourse markers as proxies, can be utilized as salient indicators for paradigmatic relations in multiple languages, outperforming patterns in terms of recall and F1-score. In addition, we observe that markers and patterns provide complementary information, leading to significant classification improvements when applied in combination.

Applying a Naive Bayes Similarity Measure to Word Sense Disambiguation
Tong Wang and Graeme Hirst

We replace the overlap mechanism of the Lesk algorithm with a simple, general-purpose Naive Bayes model that measures many-to-many association between two sets of random variables. Even with simple probability estimates such as maximum likelihood, the model gains significant improvement over the Lesk algorithm on word sense disambiguation tasks. With additional lexical knowledge from Word-Net, performance is further improved to surpass the state-of-the-art results.

Fast Easy Unsupervised Domain Adaptation with Marginalized Structured Dropout
Yi Yang and Jacob Eisenstein

Unsupervised domain adaptation often relies on transforming the instance representation. However, most such approaches are designed for bag-of-words models, and ignore the structured features present in many problems in NLP. We propose a new technique called marginalized structured dropout, which exploits feature structure to obtain a remarkably simple and efficient feature projection. Applied to the task of fine-grained part-of-speech tagging on a dataset of historical Portuguese, marginalized structured dropout yields state-of-the-art accuracy while increasing speed by more than an order-of-magnitude over previous work.

Improving Lexical Embeddings with Semantic Knowledge
Mo Yu and Mark Dredze

Word embeddings learned on unlabeled data are a popular tool in semantics, but may not capture the desired semantics. We propose a new learning objective that incorporates both a neural language model objective (Mikolov et al., 2013b) and prior knowledge from semantic resources to learn improved lexical semantic embed-dings. We demonstrate that our embeddings improve over those learned solely on raw text in three settings: language modeling, measuring semantic similarity, and predicting human judgements.

Optimizing Segmentation Strategies for Simultaneous Speech Translation
Yusuke Oda, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura

In this paper, we propose new algorithms for learning segmentation strategies for simultaneous speech translation. In contrast to previously proposed heuristic methods, our method finds a segmentation that directly maximizes the performance of the machine translation system. We describe two methods based on greedy search and dynamic programming that search for the optimal segmentation strategy. An experimental evaluation finds that our algorithm is able to segment the input two to three times more frequently than conventional methods in terms of number of words, while maintaining the same score of automatic evaluation.

A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation
Taku Kudo, Hiroshi Ichikawa, and Hideto Kazawa

We present a simple joint inference of deep case analysis and zero subject generation for the pre-ordering in Japanese-to-English machine translation. The detection of subjects and objects from Japanese sentences is more difficult than that from English, while it is the key process to generate correct English word orders.

In addition, subjects are often omitted in Japanese when they are inferable from the context.We propose a new Japanese deep syntactic parser that consists of pointwise probabilistic models and a global inference with linguistic constraints. We applied our new deep parser to pre-ordering in Japanese-to-English SMT system and show substantial improvements in automatic evaluations.

A Hybrid Approach to Skeleton-based Translation
Tong Xiao, Jingbo Zhu, and Chunliang Zhang

In this paper we explicitly consider sentence skeleton information for Machine Translation (MT). The basic idea is that we translate the key elements of the input sentence using a skeleton translation model, and then cover the remain segments using a full translation model. We apply our approach to a state-of-the-art phrase-based system and demonstrate very promising BLEU improvements and TER reductions on the NIST Chinese-English MT evaluation data.

Effective Selection of Translation Model Training Data
Le Liu, Yu Hong, Hao Liu, Xing Wang, and Jianmin Yao

Data selection has been demonstrated to be an effective approach to addressing the lack of high-quality bitext for statistical machine translation in the domain of interest. Most current data selection methods solely use language models trained on a small scale in-domain data to select domain-relevant sentence pairs from general-domain parallel corpus. By contrast, we argue that the relevance between a sentence pair and target domain can be better evaluated by the combination of language model and translation model. In this paper, we study and experiment with novel methods that apply translation models into domain-relevant data selection. The results show that our methods outperform previous methods. When the selected sentence pairs are evaluated on an end-to-end MT task, our methods can increase the translation performance by 3 BLEU points.

Refinements to Interactive Translation Prediction Based on Search Graphs
Philipp Koehn, Chara Tsoukala, and Hervé Saint-Amand

We propose a number of refinements to the canonical approach to interactive translation prediction. By more permissive matching criteria, placing emphasis on matching the last word of the user prefix, and dealing with predictions to partially typed words, we observe gains in both word prediction accuracy (+5.4%) and letter prediction accuracy (+9.3%).

Cross-lingual Model Transfer Using Feature Representation Projection
Mikhail Kozhevnikov and Ivan Titov

We propose a novel approach to cross-lingual model transfer based on feature representation projection. First, a compact feature representation relevant for the task in question is constructed for either language independently and then the mapping between the two representations is determined using parallel data. The target instance can then be mapped into the source-side feature representation using the derived mapping and handled directly by the source-side model. This approach displays competitive performance on model transfer for semantic role labeling when compared to direct model transfer and annotation projection and suggests interesting directions for further research.

Cross-language and Cross-encyclopedia Article Linking Using Mixed-language Topic Model and Hypernym Translation
Yu-Chun Wang, Chun-Kai Wu, and Richard Tzong-Han Tsai

Creating cross-language article links among different online encyclopedias is now an important task in the unification of multilingual knowledge bases. In this paper, we propose a cross-language article linking method using a mixed-language topic model and hypernym translation features based on an SVM model to link English Wikipedia and Chinese Baidu Baike, the most widely used Wiki-like encyclopedia in China. To evaluate our approach, we compile a data set from the top 500 Baidu Baike articles and their corresponding English Wiki articles. The evaluation results show that our approach achieves 80.95% in MRR and 87.46% in recall. Our method does not heavily depend on linguistic characteristics and can be easily extended to generate cross-language article links among different online encyclopedias in other languages.

Nonparametric Method for Data-driven Image Captioning
Rebecca Mason and Eugene Charniak

We present a nonparametric density estimation technique for image caption generation. Data-driven matching methods have shown to be effective for a variety of complex problems in Computer Vision. These methods reduce an inference problem for an unknown image to finding an existing labeled image which is semantically

similar. However, related approaches for image caption generation (Ordonez et al., 2011; Kuznetsova et al., 2012) are hampered by noisy estimations of visual content and poor alignment between images and human-written captions. Our work addresses this challenge by estimating a word frequency representation of the visual content of a query image. This allows us to cast caption generation as an extractive summarization problem. Our model strongly outperforms two state-of-the-art caption extraction systems according to human judgments of caption relevance.

Improved Correction Detection in Revised ESL Sentences
Huichao Xue and Rebecca Hwa

This work explores methods of automatically detecting corrections of individual mistakes in sentence revisions for ESL students. We have trained a classifier that specializes in determining whether consecutive basic-edits (word insertions, deletions, substitutions) address the same mistake. Experimental result shows that the proposed system achieves an F1 -score of 81% on correction detection and 66% for the overall system, out-performing the baseline by a large margin.

Unsupervised Alignment of Privacy Policies using Hidden Markov Models
Rohan Ramanath, Fei Liu, Norman Sadeh, and Noah A. Smith

To support empirical study of online privacy policies, as well as tools for users with privacy concerns, we consider the problem of aligning sections of a thousand policy documents, based on the issues they address. We apply an unsupervised HMM; in two new (and reusable) evaluations, we find the approach more effective than clustering and topic models.

Enriching Cold Start Personalized Language Model Using Social Network Information
Yu-Yang Huang, Rui Yan, Tsung-Ting Kuo, and Shou-De Lin

We introduce a generalized framework to enrich the personalized language models for cold start users. The cold start problem is solved with content written by friends on social network services. Our framework consists of a mixture language model, whose mixture weights are estimated with a factor graph. The factor graph is used to incorporate prior knowledge and heuristics to identify the most appropriate weights. The intrinsic and extrinsic experiments show significant improvement on cold start users.

Automatic Labelling of Topic Models Learned from Twitter by Summarisation
Amparo Elizabeth Cano Basave, Yulan He, and Ruifeng Xu

Latent topics derived by topic models such as Latent Dirichlet Allocation (LDA) are the result of hidden thematic structures which provide further insights into the data. The automatic labelling of such topics derived from social media poses however new challenges since topics may characterise novel events happening in the real world. Existing automatic topic labelling approaches which depend on external knowledge sources become less applicable here since relevant articles/concepts of the extracted topics may not exist in external sources. In this paper we propose to address the problem of automatic labelling of latent topics learned from Twitter as a summarisation problem. We introduce a framework which apply summarisation algorithms to generate topic labels. These algorithms are independent of external sources and only rely on the identification of dominant terms in documents reletated to the latent topic. We compare the efficiency of existing state of the art summarisation algorithms. Our results suggest that summarisation algorithms generate better topic labels which capture event-related context compared to the top-n terms returned by LDA.

Stochastic Contextual Edit Distance and Probabilistic FSTs
Ryan Cotterell, Nanyun Peng, and Jason Eisner

String similarity is most often measured by weighted or unweighted edit distance d(x,y). Ristad and Yianilos (1998) defined *stochastic* edit distance – a probability distribution p(y | x) whose parameters can be trained from data. We generalize this so that the probability of choosing each edit operation can depend on contextual features. We show how to construct and train a probabilistic finite-state transducer that computes our contextual stochastic edit distance. To illustrate the improvement from conditioning on context, we model typos found in social media text.

Labelling Topics using Unsupervised Graph-based Methods
Nikolaos Aletras and Mark Stevenson

This paper introduces an unsupervised graph-based method that selects textual labels for automatically generated topics. Our approach uses the topic keywords to query a search engine and generate a graph from the words contained in the results. PageRank is then used to weigh the words in the graph and score the candidate

labels. The state-of-the-art method for this task is supervised. Evaluation on a standard data set shows that the performance of our approach is consistently superior to previously reported methods.

Training a Korean SRL System with Rich Morphological Features
Young-Bum Kim, Heemoon Chae, Benjamin Snyder, and Yu-Seop Kim

In this paper we introduce a semantic role labeler for Korean, an agglutinative language with rich morphology. First, we create a novel training source by semantically annotating a Korean corpus containing fine-grained morphological and syntactic information. We then develop a supervised SRL model by leveraging morphological features of Korean that tend to correspond with semantic roles. Our model also employs a variety of latent morpheme representations induced from a larger body of unannotated Korean text. These elements lead to state-of-the-art performance of 81.07% labeled F1, representing the best SRL performance reported to date for an agglutinative language.

Semantic Parsing for Single-Relation Question Answering
Wen-tau Yih, Xiaodong He, and Christopher Meek

We develop a semantic parsing framework based on semantic similarity for open domain question answering (QA). We focus on single-relation questions and decompose each question into an entity mention and a relation pattern. Using convolutional neural network models, we measure the similarity of entity mentions with entities in the knowledge base (KB) and the similarity of relation patterns and relations in the KB. We score relational triples in the KB using these measures and select the top scoring relational triple to answer the question. When evaluated on an open-domain QA task, our method achieves higher precision across different recall points compared to the previous approach, and can improve F1 by 7 points.

On WordNet Semantic Classes and Dependency Parsing
Kepa Bengoetxea, Eneko Agirre, Joakim Nivre, Yue Zhang, and Koldo Gojenola

This paper presents experiments with WordNet semantic classes to improve dependency parsing. We study the effect of semantic classes in three dependency parsers, using two types of constituencyto- dependency conversions of the English Penn Treebank. Overall, we can say that the improvements are small and not significant using automatic POS tags, contrary to previously published results using gold POS tags (Agirre et al., 2011). In addition, we explore parser combinations, showing that the semantically enhanced parsers yield a small significant gain only on the more semantically oriented LTH treebank conversion.

Enforcing Structural Diversity in Cube-pruned Dependency Parsing
Hao Zhang and Ryan McDonald

In this paper we extend the cube-pruned dependency parsing framework of Zhang et al. (2012; 2013) by forcing inference to maintain both label and structural ambiguity. The resulting parser achieves state-of-the-art accuracies, in particular on datasets with a large set of dependency labels.

The Penn Parsed Corpus of Modern British English: First Parsing Results and Analysis
Seth Kulick, Anthony Kroch, and Beatrice Santorini

This paper presents the first results on parsing the Penn Parsed Corpus of Modern British English (PPCMBE), a million-word historical treebank with an annotation style similar to that of the Penn Treebank (PTB). We describe key features of the PPCMBE annotation style that differ from the PTB, and present some experiments with tree transformations to better compare the results to the PTB. First steps in parser analysis focus on problematic structures created by the parser.

Parser Evaluation Using Derivation Trees: A Complement to evalb
Seth Kulick, Ann Bies, Justin Mott, Anthony Kroch, Beatrice Santorini, and Mark Liberman

This paper introduces a new technique for phrase-structure parser analysis, categorizing possible treebank structures by integrating regular expressions into derivation trees. We analyze the performance of the Berkeley parser on OntoNotes WSJ and the English Web Treebank. This provides some insight into the evalb scores, and the problem of domain adaptation with the web data. We also analyze a "test-on-train" dataset, showing a wide variance in how the parser is generalizing from different structures in the training material.

Demonstrations

DKPro TC: A Java-based Framework for Supervised Learning Experiments on Textual Data
Johannes Daxenberger, Oliver Ferschke, Iryna Gurevych, and Torsten Zesch

We present DKPro TC, a framework for supervised learning experiments on textual data. The main goal of DKPro TC is to enable researchers to focus on the actual research task behind the learning problem and let the framework handle the rest. It enables rapid prototyping of experiments by relying on an easy-to-use workflow engine and standardized document preprocessing based on the Apache Unstructured Information Management Architecture (Ferrucci and Lally, 2004). It ships with standard feature extraction modules, while at the same time allowing the user to add customized extractors. The extensive reporting and logging facilities make DKPro TC experiments fully replicable.

WoSIT: A Word Sense Induction Toolkit for Search Result Clustering and Diversification
Daniele Vannella, Tiziano Flati, and Roberto Navigli

In this demonstration we present WoSIT, an API for Word Sense Induction (WSI) algorithms. The toolkit provides implementations of existing graph-based WSI algorithms, but can also be extended with new algorithms. The main mission of WoSIT is to provide a framework for the extrinsic evaluation of WSI algorithms, also within end-user applications such as Web search result clustering and diversification.

A Rule-Augmented Statistical Phrase-based Translation System
Cong Duy Vu Hoang, AiTi Aw, and Nhung T. H. Nguyen

Interactive or Incremental Statistical Machine Translation (IMT) aims to provide a mechanism that allows the statistical models involved in the translation process to be incrementally updated and improved. The source of knowledge normally comes from users who either post-edit the entire translation or just provide the translations for wrongly translated domain-specific terminologies. The majority of the existing work on IMT uses batch learning paradigm which does not allow translation system to make use of the new input instantaneously. We introduce an adaptive MT framework with a Rule Definition Language (RDL) for users to amend MT results through translation rules or patterns. Experimental results show that our system acknowledges user feedback via RDL which improves the translations of the baseline system on three test sets for Vietnamese to English translation.

KyotoEBMT: An Example-Based Dependency-to-Dependency Translation Framework
John Richardson, Fabien Cromières, Toshiaki Nakazawa, and Sadao Kurohashi

This paper introduces the KyotoEBMT Example-Based Machine Translation framework. Our system uses a tree-to-tree approach, employing syntactic dependency analysis for both source and target languages in an attempt to preserve non-local structure. The effectiveness of our system is maximized with online example matching and a flexible decoder. Evaluation demonstrates BLEU scores competitive with state-of-the-art SMT systems such as Moses. The current implementation is intended to be released as open-source in the near future.

kLogNLP: Graph Kernel–based Relational Learning of Natural Language
Mathias Verbeke, Paolo Frasconi, Kurt De Grave, Fabrizio Costa, and Luc De Raedt

kLog is a framework for kernel-based learning that has already proven successful in solving a number of relational tasks in natural language processing. In this paper, we present kLogNLP, a natural language processing module for kLog. This module enriches kLog with NLP-specific preprocessors, enabling the use of existing libraries and toolkits within an elegant and powerful declarative machine learning framework. The resulting relational model of the domain can be extended by specifying additional relational features in a declarative way using a logic programming language. This declarative approach offers a more flexible way of experimentation and to insert domain knowledge.

Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno
Seid Muhie Yimam, Chris Biemann, Richard Eckart de Castilho, and Iryna Gurevych

In this paper, we present a flexible approach to the efficient and exhaustive manual annotation of text documents. For this purpose, we extend WebAnno (Yimam et al., 2013) an open-source web-based annotation tool.1 While it was previously limited to specific annotation layers, our extension allows adding and configuring an arbitrary number of layers through a web-based UI. These layers can be annotated separately or

simultaneously, and support most types of linguistic annotations such as spans, semantic classes, dependency relations, lexical chains, and morphology. Further, we tightly integrate a generic machine learning component for automatic annotation suggestions of span annotations. In two case studies, we show that automatic annotation suggestions, combined with our split-pane UI concept, significantly reduces annotation time.

Web Information Mining and Decision Support Platform for the Modern Service Industry
Binyang Li, Lanjun Zhou, Zhongyu Wei, Kam-fai Wong, Ruifeng Xu, and Yunqing Xia

This demonstration presents an intelligent information platform MODEST. MODEST will provide enterprises with the services of retrieving news from websites, extracting commercial information, exploring customers' opinions, and analyzing collaborative/competitive social networks. In this way, enterprises can improve the competitive abilities and facilitate potential collaboration activities. At the meanwhile, MODEST can also help governments to acquire information about one single company or the entire board timely, and make prompt strategies for better support. Currently, MODEST is applied to the pillar industries of Hong Kong, including innovative finance, modem logistics, information technology, etc.

FAdR: A System for Recognizing False Online Advertisements
Yi-jie Tang and Hsin-Hsi Chen

More and more product information, including advertisements and user reviews, are presented to Internet users nowadays. Some of the information is false, misleading or overstated, which can cause seriousness and needs to be identified. Authorities, advertisers, website owners and consumers all have the needs to detect such statements. In this paper, we propose a False Advertisements Recognition system called FAdR by using one-class and binary classification models. Illegal advertising lists made public by a government and product descriptions from a shopping website are obtained for training and testing. The results show that the binary SVM models can achieve the highest performance when unigrams with the weighting of log relative frequency ratios are used as features. Comparatively, the benefit of the one-class classification models is the adjustable rejection rate parameter, which can be changed to suit different applications. Verb phrases more likely to introduce overstated information are obtained by mining the datasets. These phrases help find problematic wordings in the advertising texts.

lex4all: A language-independent tool for building and evaluating pronunciation lexicons for small-vocabulary speech recognition
Anjana Vakil, Max Paulus, Alexis Palmer, and Michaela Regneri

This paper describes lex4all, an open-source PC application for the generation and evaluation of pronunciation lexicons in any language. With just a few minutes of recorded audio and no expert knowledge of linguistics or speech technology, individuals or organizations seeking to create speech-driven applications in low-resource languages can build lexicons enabling the recognition of small vocabularies (up to 100 terms, roughly) in the target language using an existing recognition engine designed for a high-resource source language (e.g. English). To build such lexicons, we employ an existing method for cross-language phoneme-mapping. The application also offers a built-in audio recorder that facilitates data collection, a significantly faster implementation of the phoneme-mapping technique, and an evaluation module that expedites research on small-vocabulary speech recognition for low-resource languages.

Enhanced Search with Wildcards and Morphological Inflections in the Google Books Ngram Viewer
Jason Mann, David Zhang, Lu Yang, Dipanjan Das, and Slav Petrov

We present a new version of the Google Books Ngram Viewer, which plots the frequency of words and phrases over the last five centuries; its data encompasses 6% of the world's published books. The new Viewer adds three features for more powerful search: wildcards, morphological inflections, and capitalization. These additions allow the discovery of patterns that were previously difficult to find and further facilitate the study of linguistic trends in printed text.

Simplified Dependency Annotations with GFL-Web
Michael T. Mordowanec, Nathan Schneider, Chris Dyer, and Noah A. Smith

We present GFL-Web, a web-based interface for syntactic dependency annotation with the lightweight FUDG/GFL formalism. Syntactic attachments are specified in GFL notation and visualized as a graph. A one-day pilot of this workflow with 26 annotators established that even novices were, with a bit of training, able to rapidly annotate the syntax of English Twitter messages.

ACL Social Event at the National Aquarium

©Public Domain

Tuesday, June 24, 2014, 7:30pm – 10:00pm

National Aquarium in Baltimore
501 East Pratt St
Baltimore, MD 21202

Casually walk over to the National Aquarium for desserts, coffee, tea, and cash bar. Wander through the exhibits, network with colleagues, dance to the sounds of a local DJ, and enjoy a relaxing evening with the jellyfish and sharks.

Social event tickets are included with the main conference registration.

Main Conference: Wednesday, June 25

Overview

7:30 – 6:00	**Registration**				(Grand Foyer)
7:30 – 9:00	**Breakfast**				(Grand Foyer)
9:00 – 9:30	**Best paper session**				(Grand Ballroom V–VI)
10:15 – 10:45	**Coffee break**				(Grand Foyer)

	Session 7				
10:45 – 12:25	Multimodal NLP / Lexical Semantics	Semantics III	Machine Translation IV	NLP Applications and NLP Enabled Technology II	Sentiment Analysis II

12:25 – 1:30	**Lunch break**	
1:30 – 3:00	**ACL Business Meeting**	(Grand Ballroom V)

	Session 8				
3:00 – 4:30	NLP for the Web and Social Media II	Semantics / Information Extraction	Machine Translation V	Syntax, Parsing, and Tagging V	Multilinguality and Multimodal NLP

4:30 – 5:00	**Coffee break**	(Grand Foyer)
5:00 – 6:30	**Lifetime Achievement Award**	(Grand Ballroom V–VI)
6:30 – 7:00	**Closing Session**	(Grand Ballroom V–VI)

Best Paper Awards

Chairs: Kristina Toutanova and Hua Wu

Fast and Robust Neural Network Joint Models for Statistical Machine Translation
Jacob Devlin, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard Schwartz, and John Makhoul   9:00–9:30

Recent work has shown success in using neural network language models (NNLMs) as features in MT systems. Here, we present a novel formulation for a neural network joint model (NNJM), which augments the NNLM with a source context window. Our model is purely lexicalized and can be integrated into any MT decoder. Additionally, we present several variations of the NNJM which provide significant additive improvements. Although the model is quite simple, it yields extremely strong empirical results. On the NIST OpenMT12 Arabic-English condition, the NNJM features produce a gain of +3.0 BLEU on top of a powerful, feature-rich baseline which already includes a target-only NNLM. The NNJM features also produce a gain of +6.3 BLEU on top of a simpler baseline equivalent to Chiang's original Hiero implementation. Finally, we describe two novel techniques for overcoming the historically high cost of using NNLM-style models in MT decoding. These techniques speed up NNJM computation by a factor of 10,000x, making the model as fast as a standard back-off LM.

Low-Rank Tensors for Scoring Dependency Structures
Tao Lei, Yu Xin, Yuan Zhang, Regina Barzilay, and Tommi Jaakkola   9:30–10:00

Accurate scoring of syntactic structures such as head-modifier arcs in dependency parsing typically requires rich, high-dimensional feature representations. A small subset of such features is often selected manually. This is problematic when features lack clear linguistic meaning as in embeddings or when the information is blended across features. In this paper, we use tensors to map high-dimensional feature vectors into low dimensional representations. We explicitly maintain the parameters as a low-rank tensor to obtain low dimensional representations of words in their syntactic roles, and to leverage modularity in the tensor for easy training with online algorithms. Our parser consistently outperforms the Turbo and MST parsers across 14 different languages. We also obtain the best published UAS results on 5 languages.

ACL Business Meeting

Date: Wednesday, June 25, 2014
Time: 1:30–3:00 PM
Venue: Grand Ballroom V

All ACL members are encouraged to participate in the business meeting. The agenda includes updates from the ACL executive committee, presentations about upcoming conferences, and a discussion (hustings) between the two candidates for vice-president-elect: Jan Hajič and Joakim Nivre, moderated by Bonnie Webber.

Parallel Session 7 Overview - Wednesday, June 25, 2014

	Track A	**Track B**	**Track C**	**Track D**	**Track E**
	Multimodal NLP / Lexical Semantics	*Semantics III*	*Machine Translation IV*	*NLP Applications and NLP Enabled Technology II*	*Sentiment Analysis II*
	Grand V	Grand VI	Harborside A/B	Harborside C	Harborside D/E
10:45	[TACL] Back to Basics for Monolingual Alignment: Exploiting Word Similarity and Contextual Evidence *Sultan, Bethard, and Sumner*	Semantic Parsing via Paraphrasing *Berant and Liang*	A Sense-Based Translation Model for Statistical Machine Translation *Xiong and Zhang*	Predicting Instructor's Intervention in MOOC forums *Chaturvedi, Goldwasser, and Daumé III*	[TACL] Senti-LSSVM: Sentiment-Oriented Multi-Relation Extraction with Latent Structural SVM *Qu, Zhang, Wang, Jiang, Gemulla, and Weikum*
11:10	CoSimRank: A Flexible & Efficient Graph-Theoretic Similarity Measure *Rothe and Schütze*	A Discriminative Graph-Based Parser for the Abstract Meaning Representation *Flanigan, Thomson, Carbonell, Dyer, and Smith*	Recurrent Neural Networks for Word Alignment Model *Tamura, Watanabe, and Sumita*	A Joint Graph Model for Pinyin-to-Chinese Conversion with Typo Correction *Jia and Zhao*	ConnotationWordNet: Learning Connotation over the Word+Sense Network *Kang, Feng, Akoglu, and Choi*
11:35	Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world *Lazaridou, Bruni, and Baroni*	Context-dependent Semantic Parsing for Time Expressions *Lee, Artzi, Dodge, and Zettlemoyer*	A Constrained Viterbi Relaxation for Bidirectional Word Alignment *Chang, Rush, DeNero, and Collins*	Smart Selection *Pantel, Gamon, and Fuxman*	Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification *Tang, Wei, Yang, Zhou, Liu, and Qin*
12:00	[TACL] From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions *Young, Lai, Hodosh, and Hockenmaier*	Semantic Frame Identification with Distributed Word Representations *Hermann, Das, Weston, and Ganchev*	A Recursive Recurrent Neural Network for Statistical Machine Translation *Liu, Yang, Li, and Zhou*	Modeling Prompt Adherence in Student Essays *Persing and Ng*	Towards a General Rule for Identifying Deceptive Opinion Spam *Li, Ott, Cardie, and Hovy*

Parallel Session 7

Session 7A: Multimodal NLP / Lexical Semantics
Grand V   *Chair: Richard Socher*

[TACL] Back to Basics for Monolingual Alignment: Exploiting Word Similarity and Contextual Evidence
Md Arafat Sultan, Steven Bethard, and Tamara Sumner   10:45–11:10

We present a simple, easy-to-replicate monolingual aligner that demonstrates state-of-the-art performance while relying on almost no supervision and a very small number of external resources. Based on the hypothesis that words with similar meanings represent potential pairs for alignment if located in similar contexts, we propose a system that operates by finding such pairs.In two intrinsic evaluations on alignment test data, our system achieves F_1 scores of 88–92%, demonstrating 1–3% absolute improvement over the previous best system. Moreover, in two extrinsic evaluations our aligner outperforms existing aligners, and even a naive application of the aligner approaches state-of-the-art performance at each extrinsic task.

CoSimRank: A Flexible & Efficient Graph-Theoretic Similarity Measure
Sascha Rothe and Hinrich Schütze   11:10–11:35

We present CoSimRank, a new graphtheoretic similarity measure that is efficient because it can compute a single node similarity without having to compute the similarities of the entire graph. We present equivalent formalizations that show CoSimRank's close relationship to Personalized PageRank and SimRank and also show how we can take advantage of fast matrix multiplication algorithms to compute CoSimRank. Another advantage of CoSimRank is that it can be flexibly extended from basic node-node similarity to several other graph-theoretic similarity measures. In an experimental evaluation on the tasks of synonym extraction and bilingual lexicon extraction, CoSimRank is faster or more accurate than previous approaches.

Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world
Angeliki Lazaridou, Elia Bruni, and Marco Baroni   11:35–12:00

Following up on recent work on estab- lishing a mapping between vector-based semantic embeddings of words and the visual representations of the correspond- ing objects from natural images, we first present a simple approach to cross-modal vector-based semantics for the task of zero-shot learning, in which an image of a previously unseen object is mapped to a linguistic representation denoting its word. We then introduce fast mapping, a challenging and more cognitively plausi- ble variant of the zero-shot task, in which the learner is exposed to new objects and the corresponding words in very limited linguistic contexts. By combining prior linguistic and visual knowledge acquired about words and their objects, as well as exploiting the limited new evidence avail- able, the learner must learn to associate new objects with words. Our results on this task pave the way to realistic simula- tions of how children or robots could use existing knowledge to bootstrap grounded semantic knowledge about new concepts.

[TACL] From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions
Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier   12:00–12:25

We propose to use the visual denotations of linguistic expressions (i.e. the set of images they describe) to define novel denotational similarity metrics, which we show to be at least as beneficial as distributional similarities for two tasks that require semantic inference. To compute these denotational similarities, we construct a denotation graph, i.e. a subsumption hierarchy over constituents and their denotations, based on a large corpus of 30K images and 150K descriptive captions.

Session 7B: Semantics III
Grand VI   *Chair: Shiqi Zhao*

Semantic Parsing via Paraphrasing
Jonathan Berant and Percy Liang   10:45–11:10
A central challenge in semantic parsing is handling the myriad ways in which knowledge base predicates can be expressed. Traditionally, semantic parsers are trained primarily from text paired with knowledge base information. Our goal is to exploit the much larger amounts of raw text not tied to any knowledge base. In this paper, we turn semantic parsing on its head. Given an input utterance, we first use a simple method to deterministically generate a set of candidate logical forms with a canonical realization in natural language for each. Then, we use a paraphrase model to choose the realization that best paraphrases the input, and output the corresponding logical form. We present two simple paraphrase models, an association model and a vector space model, and train them jointly from question-answer pairs. Our system PARASEMPRE improves state-of-the-art accuracies on two recently released question-answering datasets.

A Discriminative Graph-Based Parser for the Abstract Meaning Representation
Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith   11:10–11:35
Abstract meaning representation is a semantic formalism for which a growing set of annotated examples is available. We introduce the first approach to parsing sentences into this representation, providing a strong baseline for future improvement. The method is based on a novel algorithm for finding a maximum spanning, connected subgraph, embedded within a Lagrangian relaxation of a hard optimization problem that imposes linguistically inspired constraints. Our approach is described in the general framework of structured prediction, allowing future incorporation of additional features and constraints, and may extend to other formalisms as well.

Context-dependent Semantic Parsing for Time Expressions
Kenton Lee, Yoav Artzi, Jesse Dodge, and Luke Zettlemoyer   11:35–12:00
We present an approach for learning context-dependent semantic parsers to identify and interpret time expressions. We use a Combinatory Categorial Grammar to construct compositional meaning representations, while considering contextual cues, such as the document creation time and the tense of the governing verb, to compute the final time values. Experiments on benchmark datasets show that our approach outperforms previous state-of-the-art systems, with error reductions of 13% to 21% in end-to-end performance.

Semantic Frame Identification with Distributed Word Representations
Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev   12:00–12:25
We present a novel technique for semantic frame identification using distributed representations of predicates and their syntactic context; this technique leverages automatic syntactic parses and a generic set of word embeddings. Given labeled data annotated with frame-semantic parses, we learn a model that projects the set of word representations for the syntactic context around a predicate to a low dimensional representation. The latter is used for semantic frame identification; with a standard argument identification method inspired by prior work, we achieve state-of-the-art results on FrameNet-style frame-semantic analysis. Additionally, we report strong results on PropBank-style semantic role labeling in comparison to prior work.

Session 7C: Machine Translation IV

Harborside A/B   *Chair: David Chiang*

A Sense-Based Translation Model for Statistical Machine Translation

Deyi Xiong and Min Zhang   10:45–11:10

The sense in which a word is used determines the translation of the word. In this paper, we propose a sense-based translation model to integrate word senses into statistical machine translation. We build a broad-coverage sense tagger based on a nonparametric Bayesian topic model that automatically learns sense clusters for words in the source language. The proposed sense-based translation model enables the decoder to select appropriate translations for source words according to the inferred senses for these words using maximum entropy classifiers. Our method is significantly different from previous word sense disambiguation reformulated for machine translation in that the latter neglects word senses in nature. We test the effectiveness of the proposed sense-based translation model on a large-scale Chinese-to-English translation task. Results show that the proposed model substantially outperforms not only the baseline but also the previous reformulated word sense disambiguation.

Recurrent Neural Networks for Word Alignment Model

Akihiro Tamura, Taro Watanabe, and Eiichiro Sumita   11:10–11:35

This study proposes a word alignment model based on a recurrent neural network (RNN), in which an unlimited alignment history is represented by recurrently connected hidden layers. We perform unsupervised learning using noise-contrastive estimation, which utilizes artificially generated negative samples. Our alignment model is directional, similar to the generative IBM models. To overcome this limitation, we encourage agreement between the two directional models by introducing a penalty function that ensures word embedding consistency across two directional models during training. The RNN-based model outperforms the feed-forward neural network-based model (Yang et al., 2013) as well as the IBM Model 4 under Japanese-English and French-English word alignment tasks, and achieves comparable translation performance to those baselines for Japanese-English and Chinese-English translation tasks.

A Constrained Viterbi Relaxation for Bidirectional Word Alignment

Yin-Wen Chang, Alexander M. Rush, John DeNero, and Michael Collins   11:35–12:00

Bidirectional models of word alignment are an appealing alternative to post-hoc combinations of directional word aligners. Unfortunately, most formulations are NP-Hard to solve, and a previous attempt to use a relaxation-based decoder yielded few exact solutions (6

A Recursive Recurrent Neural Network for Statistical Machine Translation

Shujie Liu, Nan Yang, Mu Li, and Ming Zhou   12:00–12:25

In this paper, we propose a novel recursive recurrent neural network (R2NN) to model the end-to-end decoding process for statistical machine translation. R2NN is a combination of recursive neural network and recurrent neural network, and in turn integrates their respective capabilities: (1) new information can be used to generate the next hidden state, like recurrent neural networks, so that language model and translation model can be integrated naturally; (2) a tree structure can be built, as recursive neural networks, so as to generate the translation candidates in a bottom up manner. A semi-supervised training approach is proposed to train the parameters, and the phrase pair embedding is explored to model translation confidence directly. Experiments on a Chinese to English translation task show that our proposed R2NN can outperform the state-of-the-art baseline by about 1.5 points in BLEU.

Session 7D: NLP Applications and NLP Enabled Technology II
Harborside C   *Chair: Brian Roark*

Predicting Instructor's Intervention in MOOC forums
Snigdha Chaturvedi, Dan Goldwasser, and Hal Daumé III   10:45–11:10

Instructor intervention in student discussion forums is a vital component in Massive Open Online Courses (MOOCs), where personalized interaction is limited. This paper introduces the problem of predicting instructor interventions in MOOC forums. We propose several prediction models designed to capture unique aspects of MOOCs, combining course information, forum structure and posts content. Our models abstract contents of individual posts of threads using latent categories, learned jointly with the binary intervention prediction problem. Experiments over data from two Coursera MOOCs demonstrate that incorporating the structure of threads into the learning problem leads to better predictive performance.

A Joint Graph Model for Pinyin-to-Chinese Conversion with Typo Correction
Zhongye Jia and Hai Zhao   11:10–11:35

It is very import for Chinese language processing with the aid of an efficient input method engine (IME), of which pinyin-to-Chinese (PTC) conversion is the core part. Meanwhile, though typos are inevitable during user pinyin inputting, existing IMEs paid little attention to such big inconvenience. In this paper, motivated by a key equivalence of two decoding algorithms, we propose a joint graph model to globally optimize PTC and typo correction for IME. The evaluation results show that the proposed method outperforms both existing academic and commercial IMEs.

Smart Selection
Patrick Pantel, Michael Gamon, and Ariel Fuxman   11:35–12:00

Natural touch interfaces, common now in devices such as tablets and smartphones, make it cumbersome for users to select text. There is a need for a new text selection paradigm that goes beyond the high acuity selection-by-mouse that we have relied on for decades. In this paper, we introduce such a paradigm, called "Smart Selection", which aims to recover a user's intended text selection from her touch input. We model the problem using an ensemble learning approach, which leverages multiple linguistic analysis techniques combined with information from a knowledge base and a Web graph. We collect a dataset of true intended user selections and simulated user touches via a large-scale crowdsourcing task, which we release to the academic community. We show that our model effectively addresses the smart selection task and significantly outperforms various baselines and standalone linguistic analysis techniques.

Modeling Prompt Adherence in Student Essays
Isaac Persing and Vincent Ng   12:00–12:25

Recently, researchers have begun exploring methods of scoring student essays with respect to particular dimensions of quality such as coherence, technical errors, and prompt adherence. The work on modeling prompt adherence, however, has been focused on whether individual sentences adhere to the prompt. We present a new annotated corpus of essay-level prompt adherence scores and propose a learning-based approach to scoring essays along the prompt adherence dimension. Our approach makes use of novel features and significantly outperforms a baseline prompt adherence scoring system yielding relative error reductions of up to 21.4%.

Session 7E: Sentiment Analysis II
Harborside D/E   *Chair: Janyce Wiebe*

[TACL] Senti-LSSVM: Sentiment-Oriented Multi-Relation Extraction with Latent Structural SVM
Lizhen Qu, Yi Zhang, Rui Wang, Lili Jiang, Rainer Gemulla, and Gerhard Weikum   10:45–11:10

Sentiment-oriented multi-relation extraction from user-generated web documents is of great importance for both producers and consumers. Unlike the previous work, we formulate this extraction task as a learning problem with structured outputs and design the corresponding inference as an integer linear program. Our newly proposed model based on latent structural SVM can learn from the training corpora, which does not contain explicit annotations of sentiment-bearing expressions, and it can simultaneously recognize both binary (opinion) and ternary (comparison) relations with regard to entity mentions of interest. The empirical evaluation shows that our approach significantly outperforms the state-of-the-art baselines across domains (camera and movie) and across genres (reviews and forum posts). The gold standard corpus will also be a valuable resource for the community.

ConnotationWordNet: Learning Connotation over the Word+Sense Network
Jun Seok Kang, Song Feng, Leman Akoglu, and Yejin Choi   11:10–11:35

We introduce ConnotationWordNet, a connotation lexicon over the network of words in conjunction with senses. We formulate the lexicon induction problem as collective inference over pairwise-Markov Random Fields, and present a loopy belief propagation algorithm for inference. The key aspect of our method is that it is the first unified approach that assigns the polarity of both word- and sense-level connotations, exploiting the innate bipartite graph structure encoded in WordNet. We present comprehensive evaluation to demonstrate the quality and utility of the resulting lexicon in comparison to existing connotation and sentiment lexicons.

Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification
Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin   11:35–12:00

We present a method that learns word embedding for Twitter sentiment classification in this paper. Most existing algorithms for learning continuous word representations typically model only the syntactic context of words but ignore the sentiment of text. This is problematic for sentiment analysis as they usually map words with similar syntactic context but opposite sentiment polarity, such as good and bad, to neighboring word vectors. We address this issue by learning sentiment-specific word embedding (SSWE), which encodes sentiment information in the continuous representation of words. Specifically, we develop three neural network architectures to effectively incorporate the supervision from sentiment polarity of text (e.g. sentences or tweets) in their loss functions. To obtain large scale training corpora, we learn the sentiment-specific word embedding from massive distant-supervised tweets collected by positive and negative emoticons. Experiments on applying SSWE to a benchmark Twitter sentiment classification dataset in SemEval 2013 show that (1) the SSWE feature performs comparably with hand-crafted features in the top-performed system; (2) the performance is further improved by concatenating SSWE with existing feature set.

Towards a General Rule for Identifying Deceptive Opinion Spam
Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy   12:00–12:25

Consumers' purchase decisions are increasingly influenced by user-generated online re- views. Accordingly, there has been growing concern about the potential for posting deceptive opinion spam fictitious reviews that have been deliberately written to sound authentic, to deceive the readers. In this paper, we explore generalized approaches for identifying online deceptive opinion spam based on a new gold standard data, which is comprised of data from four different domains (i.e. Hotel, Restaurant, Airline, Doctor), each of which contains three types of reviews, i.e. costumer generated truthful reviews, Turker generated deceptive reviews and employee (domain-expert) generated deceptive reviews. Our approach tries to capture the general difference of language usage between deceptive and truthful reviews and we hope that it would be of help to customers when making purchase decisions, and review portal operators, such as TripAdvisor or Yelp to investigate fraudulent activity on their sites.

Parallel Session 8 Overview - Wednesday, June 25, 2014

	Track A	Track B	Track C	Track D	Track E
	NLP for the Web and Social Media II	*Semantics / Information Extraction*	*Machine Translation V*	*Syntax, Parsing, and Tagging V*	*Multilinguality and Multimodal NLP*
	Grand V	Grand VI	Harborside A/B	Harborside C	Harborside D/E
3:00	Learning Polylingual Topic Models from Code-Switched Social Media Documents *Peng, Wang, and Dredze*	Applying Grammar Induction to Text Mining *Salway and Touileb*	Empirical Study of Unsupervised Chinese Word Segmentation Methods for SMT on Large-scale Corpora *Wang, Utiyama, Finch, and Sumita*	Punctuation Processing for Projective Dependency Parsing *Ma, Zhang, and Zhu*	Distributed Representations of Geographically Situated Language *Bamman, Dyer, and Smith*
3:15	Normalizing tweets with edit scripts and recurrent neural embeddings *Chrupała*	Semantic Consistency: A Local Subspace Based Method for Distant Supervised Relation Extraction *Han and Sun*	EM Decipherment for Large Vocabularies *Nuhn and Ney*	Transforming trees into hedges and parsing with "hedgebank" grammars *Yarmohammadi, Dunlop, and Roark*	Improving Multi-Modal Representations Using Image Dispersion: Why Less is Sometimes More *Kiela, Hill, Korhonen, and Clark*
3:30	Exponential Reservoir Sampling for Streaming Language Models *Osborne, Lall, and Van Durme*	Concreteness and Subjectivity as Dimensions of Lexical Meaning *Hill and Korhonen*	XMEANT: Better semantic MT evaluation without reference translations *Lo, Beloucif, Saers, and Wu*	Incremental Predictive Parsing with TurboParser *Köhn and Menzel*	Bilingual Event Extraction: a Case Study on Trigger Type Determination *Zhu, Li, Zhou, and Xia*
3:45	A Piece of My Mind: A Sentiment Analysis Approach for Online Dispute Detection *Wang and Cardie*	Infusion of Labeled Data into Distant Supervision for Relation Extraction *Pershina, Min, Xu, and Grishman*	Sentence Level Dialect Identification for Machine Translation System Selection *Salloum, Elfardy, Alamir-Salloum, Habash, and Diab*	Tailoring Continuous Word Representations for Dependency Parsing *Bansal, Gimpel, and Livescu*	Understanding Relation Temporality of Entities *Lee and Hwang*
4:00	A Simple Bayesian Modelling Approach to Event Extraction from Twitter *Zhou, Chen, and He*	Recognizing Implied Predicate-Argument Relationships in Textual Inference *Stern and Dagan*	RNN-based Derivation Structure Prediction for SMT *Zhai, Zhang, Zhou, and Zong*	Observational Initialization of Type-Supervised Taggers *Zhang and DeNero*	Does the Phonology of L1 Show Up in L2 Texts? *Nicolai and Kondrak*
4:15	Be Appropriate and Funny: Automatic Entity Morph Encoding *Zhang, Huang, Pan, Ji, Knight, Wen, Sun, Han, and Yener*	Measuring metaphoricity *Dunn*	Hierarchical MT Training using Max-Violation Perceptron *Zhao, Huang, Mi, and Ittycheriah*	How much do word embeddings encode about syntax? *Andreas and Klein*	Cross-lingual Opinion Analysis via Negative Transfer Detection *Gui, Xu, Lu, Xu, Xu, Liu, and Wang*

Parallel Session 8

Session 8A: NLP for the Web and Social Media II
Grand V   *Chair: Alan Ritter*

Learning Polylingual Topic Models from Code-Switched Social Media Documents
Nanyun Peng, Yiming Wang, and Mark Dredze   3:00–3:15

Code-switched documents are common in social media, providing evidence for polylingual topic models to infer aligned topics across languages. We present Code-Switched LDA (csLDA), which infers language specific topic distributions based on code-switched docu- ments to facilitate multi-lingual corpus anal- ysis. We experiment on two code-switching corpora (English-Spanish Twitter data and English-Chinese Weibo data) and show that csLDA improves perplexity over LDA, and learns semantically coherent aligned topics as judged by human annotators.

Normalizing tweets with edit scripts and recurrent neural embeddings
Grzegorz Chrupała   3:15–3:30

Tweets often contain a large proportion of abbreviations, alternative spellings, novel words and other non-canonical language. These features are problematic for standard language analysis tools and it can be desirable to convert them to canonical form. We propose a novel text normalization model based on learning edit operations from labeled data while incorporating features induced from unlabeled data via character-level neural text embeddings. The text embeddings are generated using an Simple Recurrent Network. We find that enriching the feature set with text embeddings substantially lowers word error rates on an English tweet normalization dataset. Our model improves on state-of-the-art with little training data and without any lexical resources.

Exponential Reservoir Sampling for Streaming Language Models
Miles Osborne, Ashwin Lall, and Benjamin Van Durme   3:30–3:45

We show how rapidly changing textual streams such as Twitter can be modelled in fixed space. Our approach is based upon a randomised algorithm called *Exponential Reservoir Sampling*, unexplored by this community until now. Using language models over Twitter and Newswire as a testbed, our experimental results based on perplexity support the intuition that recently observed data generally outweighs that seen in the past, but that at times, the past can have valuable signals enabling better modelling of the present.

A Piece of My Mind: A Sentiment Analysis Approach for Online Dispute Detection
Lu Wang and Claire Cardie   3:45–4:00

We investigate the novel task of online dispute detection and propose a sentiment analysis solution to the problem: we aim to identify the sequence of sentence-level sentiments expressed during a discussion and to use them as features in a classifier that predicts the DISPUTE/NON-DISPUTE label for the discussion as a whole. We evaluate dispute detection approaches on a newly created corpus of Wikipedia Talk page disputes and find that classifiers that rely on our sentiment tagging features outperform those that do not. The best model achieves a very promising F1 score of 0.78 and an accuracy of 0.80.

A Simple Bayesian Modelling Approach to Event Extraction from Twitter
Deyu Zhou, Liangyu Chen, and Yulan He   4:00–4:15

With the proliferation of social media sites, social streams have proven to contain the most up-to-date information on current events. Therefore, it is crucial to extract events from the social streams such as tweets. However, it is not straightforward to adapt the existing event extraction systems since texts in social media are fragmented and noisy. In this paper we propose a simple and yet effective Bayesian model, called Latent Event Model (LEM), to extract structured representation of events from social media. LEM is fully unsupervised and does not require annotated data for training. We evaluate LEM on a Twitter corpus. Experimental results show that the proposed model achieves 83

Be Appropriate and Funny: Automatic Entity Morph Encoding
Boliang Zhang, Hongzhao Huang, Xiaoman Pan, Heng Ji, Kevin Knight, Zhen Wen, Yizhou Sun, Jiawei Han, and Bulent Yener   4:15–4:30

Internet users are keen on creating different kinds of morphs to avoid censorship, express strong sentiment or humor. For example, in Chinese social media, users often use the entity morph "XXX (Instant Noodles)" to

refer to "XXX (Zhou Yongkang)" because it shares one character "X (Kang)" with the well-known brand of instant noodles "XXX (Master Kang)". We developed a wide variety of novel approaches to automatically encode proper and interesting morphs, which can effectively pass decoding tests.

Session 8B: Semantics / Information Extraction

Grand VI   *Chair: Percy Liang*

Applying Grammar Induction to Text Mining

Andrew Salway and Samia Touileb   3:00–3:15

We report the first steps of a novel investigation into how a grammar induction algorithm can be modified and used to identify salient information structures in a corpus. The information structures are to be used as representations of semantic content for text mining purposes. We modify the learning regime of the ADIOS algorithm (Solan et al., 2005) so that text is presented as increasingly large snippets around key terms, and instances of selected structures are substituted with common identifiers in the input for subsequent iterations. The technique is applied to 1.4m blog posts about climate change which mention diverse topics and reflect multiple perspectives and different points of view. Observation of the resulting information structures suggests that they could be useful as representations of semantic content. Preliminary analysis shows that our modifications had a beneficial effect for inducing more useful structures.

Semantic Consistency: A Local Subspace Based Method for Distant Supervised Relation Extraction

Xianpei Han and Le Sun   3:15–3:30

One fundamental problem of distant supervision is the noisy training corpus problem. To resolve this problem, this paper proposes a new distant supervision method, called Semantic Consistency, which can identify reliable instances from noisy instances by in-specting whether an instance is located in a semantically consistent region. Specifically, we propose a semantic consistency model, which first models the local subspace around an instance as a sparse linear combination of training instances, then estimate the semantic consistency using the characteristics of the local subspace. Experimental results verified the effectiveness of our method.

Concreteness and Subjectivity as Dimensions of Lexical Meaning

Felix Hill and Anna Korhonen   3:30–3:45

We quantify the lexical subjectivity of adjectives using a corpus-based method, and show for the first time that it correlates with noun concreteness in large corpora. These cognitive dimensions together influence how word meanings combine, and we exploit this fact to achieve performance improvements on the semantic classification of adjective-noun pairs.

Infusion of Labeled Data into Distant Supervision for Relation Extraction

Maria Pershina, Bonan Min, Wei Xu, and Ralph Grishman   3:45–4:00

Distant supervision usually utilizes only unlabeled data and existing knowledge bases to learn relation extraction models. However, in some cases a small amount of human labeled data is available. In this paper, we demonstrate how a state-of-the-art multi-instance multi-label model can be profitably modified to make use of these reliable sentence-level labels in addition to the relation-level distant supervision from a database. Experiments show that our approach achieves a 13.5% increase in F-score and a 37% increase in area under the precision recall curve.

Recognizing Implied Predicate-Argument Relationships in Textual Inference

Asher Stern and Ido Dagan   4:00–4:15

We investigate recognizing implied predicate-argument relationships which are not explicitly expressed in syntactic structure. While prior works addressed such relationships as an extension to semantic role labeling, our work investigates them in the context of textual inference scenarios. Such scenarios provide prior information, which substantially eases the task. We provide a large and freely available evaluation dataset for our task setting, and propose methods to cope with it, while obtaining promising results in empirical evaluations.

Measuring metaphoricity

Jonathan Dunn   4:15–4:30

This paper presents the first computationally-derived scalar measurement of metaphoricity. Each input sentence is given a value between 0 and 1 which represents how metaphoric that sentence is. This measure achieves a correlation of 0.450 (Pearson's R, $p < 0.01$) with an experimental measure of metaphoricity involving human participants. While far from perfect, this scalar measure of metaphoricity allows different thresholds for metaphoricity so that metaphor identification can be fitted for specific tasks and datasets. When reduced to a binary classification evaluation using the VU Amsterdam Metaphor Corpus, the system achieves an F-Measure of 0.608, slightly lower than the comparable binary classification system's 0.638 and competitive

with existing approaches.

Session 8C: Machine Translation V

Harborside A/B   *Chair: Yanjun Ma*

Empirical Study of Unsupervised Chinese Word Segmentation Methods for SMT on Large-scale Corpora

Xiaolin Wang, Masao Utiyama, Andrew Finch, and Eiichiro Sumita   3:00–3:15

Unsupervised word segmentation (UWS) can provide domain-adaptive segmentation for statistical machine translation (SMT) without annotated data, and bilingual UWS can even optimize segmentation for alignment. Monolingual UWS approaches of explicitly modeling the probabilities of words through Dirichlet process (DP) models or Pitman-Yor process (PYP) models have achieved high accuracy, but their bilingual counterparts have only been carried out on small corpora such as basic travel expression corpus (BTEC) due to the computational complexity. This paper proposes an efficient unified PYP-based monolingual and bilingual UWS method. Experimental results show that the proposed method is comparable to supervised segmenters on the in-domain NIST OpenMT corpus, and yields a 0.96 BLEU relative increase on NTCIR PatentMT corpus which is out-of-domain.

EM Decipherment for Large Vocabularies

Malte Nuhn and Hermann Ney   3:15–3:30

This paper addresses the problem of EM based decipherment for large vocabularies. Here, decipherment is essentially a tagging problem: Every cipher token is tagged with some plaintext type. As with other tagging problems, this one can be treated as a Hidden Markov Model (HMM), only here, the vocabularies are large, so the usual $O(NV^2)$ exact EM approach is infeasible. When faced with this situation, many people turn to sampling. However, we propose to use a type of approximate EM and show that it works well. The basic idea is to collect fractional counts only over a small subset of links in the forward-backward lattice. The subset is different for each iteration of EM. One option is to use beam search to do the subsetting. The second method restricts the successor words that are looked at, for each hypothesis. It does this by consulting pre-computed tables of likely n-grams and likely substitutions.

XMEANT: Better semantic MT evaluation without reference translations

Chi-kiu Lo, Meriem Beloucif, Markus Saers, and Dekai Wu   3:30–3:45

We introduce XMEANT—a new cross-lingual version of the semantic frame based MT evaluation metric MEANT—which can correlate even more closely with human adequacy judgments than monolingual MEANT and eliminates the need for expensive human references. Previous work established that MEANT reflects translation adequacy with state-of-the-art accuracy, and optimizing MT systems against MEANT robustly improves translation quality. However, to go beyond tuning weights in the loglinear SMT model, a cross-lingual objective function that can deeply integrate semantic frame criteria into the MT training pipeline is needed. We show that cross-lingual XMEANT outperforms monolingual MEANT by (1) replacing the monolingual context vector model in MEANT with simple translation probabilities, and (2) incorporating bracketing ITG constraints.

Sentence Level Dialect Identification for Machine Translation System Selection

Wael Salloum, Heba Elfardy, Linda Alamir-Salloum, Nizar Habash, and Mona Diab   3:45–4:00

In this paper we study the use of sentence-level dialect identification in optimizing machine translation system selection when translating mixed dialect input. We test our approach on Arabic, a prototypical diglossic language; and we optimize the combination of four different machine translation systems. Our best result improves over the best single MT system baseline by 1.0% BLEU and over a strong system selection baseline by 0.6% BLEU on a blind test set.

RNN-based Derivation Structure Prediction for SMT

Feifei Zhai, Jiajun Zhang, Yu Zhou, and Chengqing Zong   4:00–4:15

In this paper, we propose a novel derivation structure prediction (DSP) model for SMT using recursive neural network (RNN). Within the model, two steps are involved: (1) phrase-pair vector representation, to learn vector representations for phrase pairs; (2) derivation structure prediction, to generate a bilingual RNN that aims to distinguish good derivation structures from bad ones. Final experimental results show that our DSP model can significantly improve the translation quality.

Hierarchical MT Training using Max-Violation Perceptron

Kai Zhao, Liang Huang, Haitao Mi, and Abraham Ittycheriah   4:15–4:30

Large-scale discriminative training has become promising for statistical machine translation by leveraging the huge training corpus; for example the recent effort in phrase-based MT (Yu et al., 2013) significantly outperforms mainstream methods that only train on small tuning sets. However, phrase-based MT suffers from limited reorderings, and thus its training can only utilize a small portion of the bitext due to the distortion limit. To address this problem, we extend Yu et al. (2013) to syntax-based MT by generalizing their latent variable "violation-fixing" perceptron from graphs to hypergraphs. Experiments confirm that our method leads to up to +1.2 B LEU improvement over mainstream methods such as MERT and PRO .

Session 8D: Syntax, Parsing, and Tagging V
Harborside C   *Chair: Mark Johnson*

Punctuation Processing for Projective Dependency Parsing
Ji Ma, Yue Zhang, and Jingbo Zhu   3:00–3:15

Modern statistical dependency parsers assign lexical heads to punctuations as well as words. Punctuation parsing errors lead to low parsing accuracy on words. In this work, we propose an alternative approach to addressing punctuation in dependency parsing. Rather than assigning lexical heads to punctuations, we treat punctuations as properties of their neighbouring words, used as features to guide the parser to build the dependency graph. Integrating our method with a arc-standard parser yields a 93.06% unlabelled attachment score, which is the best accuracy by a single-model transition-based parser reported so far.

Transforming trees into hedges and parsing with "hedgebank" grammars
Mahsa Yarmohammadi, Aaron Dunlop, and Brian Roark   3:15–3:30

Finite-state chunking and tagging methods are very fast for annotating non-hierarchical syntactic information, and are often applied in applications that do not require full syntactic analyses. Scenarios such as incremental machine translation may benefit from some degree of hierarchical syntactic analysis without requiring fully connected parses. We introduce hedge parsing as an approach to recovering constituents of length up to some maximum span L. This approach improves efficiency by bounding constituent size, and allows for efficient segmentation strategies prior to parsing. Unlike shallow parsing methods, hedge parsing yields internal hierarchical structure of phrases within its span bound. We present the approach and some initial experiments on different inference strategies.

Incremental Predictive Parsing with TurboParser
Arne Köhn and Wolfgang Menzel   3:30–3:45

Most approaches to incremental parsing either incur a degradation of accuracy or they have to postpone decisions, yielding underspecified intermediate output. We present an incremental predictive dependency parser that is fast, accurate, and largely language independent. By extending a state-of-the-art dependency parser, connected analyses for sentence prefixes are obtained, which even predict properties and the structural embedding of upcoming words. In contrast to other approaches, accuracy for complete sentence analyses does not decrease.

Tailoring Continuous Word Representations for Dependency Parsing
Mohit Bansal, Kevin Gimpel, and Karen Livescu   3:45–4:00

Word representations have proven useful for many NLP tasks, e.g., Brown clusters as features in dependency parsing (Koo et al., 2008). In this paper, we investigate the use of continuous word representations as features for dependency parsing. We compare several popular embeddings to Brown clusters, via multiple types of features, in both news and web domains. We find that all embeddings yield significant parsing gains, including some recent ones that can be trained in a fraction of the time of others. Explicitly tailoring the representations for the task leads to further improvements. Moreover, an ensemble of all representations achieves the best results, suggesting their complementarity.

Observational Initialization of Type-Supervised Taggers
Hui Zhang and John DeNero   4:00–4:15

Recent work has sparked new interest in type-supervised part-of-speech tagging, a data setting in which no labeled sentences are available, but the set of allowed tags is known for each word type. This paper describes observational initialization, a novel technique for initializing EM when training a type-supervised HMM tagger. Our initializer allocates probability mass to unambiguous transitions in an unlabeled corpus, generating token-level observations from type-level supervision. Experimentally, observational initialization gives state-of-the-art type-supervised tagging accuracy, providing an error reduction of 56% over uniform initialization on the Penn English Treebank.

How much do word embeddings encode about syntax?
Jacob Andreas and Dan Klein   4:15–4:30

How might a parser benefit from real-valued word embeddings? We isolate three ways in which word embeddings could be used to extend a state-of-the-art statistical parser: by relating out-of-vocabulary words to known ones, by encouraging common behavior among related in-vocabulary words, and by directly providing features for the lexicon. On small training sets, we show small but significant gains over the baseline

parser; as the training data grows, we find that these gains diminish. Our results support a hypothesis that word embeddings import syntactic information that is useful but redundant with (rather than complementary to) distinctions learned from larger treebanks in other ways.

Session 8E: Multilinguality and Multimodal NLP
Harborside D/E   *Chair: Xiaodong He*

Distributed Representations of Geographically Situated Language
David Bamman, Chris Dyer, and Noah A. Smith   3:00–3:15

We introduce a model for incorporating contextual information (such as geography) in learning vector-space representations of situated language. In contrast to approaches to multimodal representation learning that have used properties of the object being described (such as its color), our model includes information about the subject (i.e., the speaker), allowing us to learn the contours of a word's meaning that are shaped by the context in which it is uttered. In a quantitative evaluation on the task of judging geographically informed semantic similarity between representations learned from 1.1 billion words of geo-located tweets, our joint model outperforms comparable independent models that learn meaning in isolation.

Improving Multi-Modal Representations Using Image Dispersion: Why Less is Sometimes More
Douwe Kiela, Felix Hill, Anna Korhonen, and Stephen Clark   3:15–3:30

Models that learn semantic representations from both linguistic and perceptual input outperform text-only models in many contexts and better reflect human concept acquisition. However, experiments suggest that while the inclusion of perceptual input improves representations of certain concepts, it degrades the representations of others. We propose an unsupervised method to determine whether to include perceptual input for a concept, and show that it significantly improves the ability of multi-modal models to learn and represent word meanings. The method relies solely on image data, and can be applied to a variety of other NLP tasks.

Bilingual Event Extraction: a Case Study on Trigger Type Determination
Zhu Zhu, Shoushan Li, Guodong Zhou, and Rui Xia   3:30–3:45

Event extraction generally suffers from the data sparseness problem. In this paper, we address this problem by utilizing the labeled data from two different languages. As a preliminary study, we mainly focus on the subtask of trigger type determination in event extraction. To make the training data in different languages help each other, we propose a uniform text representation with bilingual features to represent the samples and handle the difficulty of locating the triggers in the translated text from both monolingual and bilingual perspectives. Empirical studies demonstrate the effectiveness of the proposed approach to bilingual classification on trigger type determination.

Understanding Relation Temporality of Entities
Taesung Lee and Seung-won Hwang   3:45–4:00

This paper demonstrates the importance of relation equivalence for entity translation pair discovery. Existing approach of understanding relation equivalence has focused on using explicit features of co-occurring entities. In this paper, we explore latent features of temporality for understanding relation equivalence, and empirically show that the explicit and latent features complement each other. Our proposed hybrid approach of using both explicit and latent features improves relation translation by 0.16 F1-score, and in turn improves entity translation by 0.02.

Does the Phonology of L1 Show Up in L2 Texts?
Garrett Nicolai and Grzegorz Kondrak   4:00–4:15

The relative frequencies of character bigrams appear to contain much information for predicting the first language (L1) of the writer of a text in another language (L2). Tsur and Rappoport (2007) interpret this fact as evidence that word choice is dictated by the phonology of L1. In order to test their hypothesis, we design an algorithm to identify the most discriminative words and the corresponding character bigrams,and perform two experiments to quantify their impact on the L1 identification task. The results strongly suggest an alternative explanation of the effectiveness of character bigrams in identifying the native language of a writer.

Cross-lingual Opinion Analysis via Negative Transfer Detection
Lin Gui, Ruifeng Xu, Qin Lu, Jun Xu, Jian Xu, Bin Liu, and Xiaolong Wang   4:15–4:30

Transfer learning method has been used in opinion analysis to make use of available language resources for other resource-poor languages. However, the cumulative class noise in transfer learning can adversely affect performance when more training data is used. In this paper, we propose a novel method in transductive transfer learning to identify noises through the detection of negative transfers. Evaluation on NLP&CC 2013 cross-lingual opinion analysis dataset shows that our approach outperformed the state-of-the-art systems. More

signifi-cantly, our system shows a monotonic increase trend in performance improvement when more training data are used.

ACL Lifetime Achievement Award

Date: Wednesday, June 25, 2014
Time: 5:00–6:30 PM
Venue: Grand Ballroom V–VI

The ACL Lifetime Achievement Award (LTA) was instituted on the occasion of the Association's 40th anniversary meeting. The award is presented for scientific achievement of both theoretical and applied nature in the field of Computational Linguistics. Currently, an ACL committee nominates and selects at most one award recipient annually, considering the originality, depth, breadth, and impact of the entire body of the nominee's work in the field. The award is a crystal trophy and the recipient is invited to give a 45-minute speech on his or her view of the development of Computational Linguistics at the annual meeting of the association. As of 2004, the speech has been subsequently published in the Association's journal, Computational Linguistics. The speech is introduced by the announcement of the award winner, whose identity is not made public until that time.

Previous winners of the distinguished award have been: Aravind Joshi (2002), Makoto Nagao (2003), Karen Spärck Jones (2004), Martin Kay (2005), Eva Hajicová (2006), Lauri Karttunen (2007), Yorick Wilks (2008), Fred Jelinek (2009), William A. Woods (2010), Eugene Charniak (2011), Charles Fillmore (2012), and Jerry Hobbs (2013).

Special Cultural Event: Thursday, June 26, 8:00 PM

ACL 2014 welcomes all attendees to an outstanding professional production of the acclaimed satirical play about language - *The Memo* - written by dissident playwright Václav Havel, the first president of the Czech Republic.

This special reprise performance of a recent sold-out production of Baltimore's award-winning Single Carrot Theater will be held in the Arellano Theater of Johns Hopkins University.

Tickets are free at the ACL registration desk to the first 100 requesting attendees, thanks to the generous full sponsorship of the Center for Language and Speech Processing at Johns Hopkins University.

The play will be preceded by a 7:15PM welcome reception and optional tour of the Johns Hopkins University Homewood campus and its Center for Language and Speech Processing, one of the two largest NLP/MT/Speech programs in the United States. Transportation details will be provided to ticket holders at the conference.

The Memo
By Václav Havel

Translated from Czech by Paul Wilson
Directed by Stephen Nunns

A comic workplace classic about linguistics and the perils of standardized communication.

Company President Mr. Gross receives a memorandum but can't read it because it's written in Ptydepe*, the newly invented language to which all correspondence must adhere. If he can't figure out what it says, he'll lose his job and certainly his mind. This satirical take on bureaucracy and office malarkey is an incisive look at 20th century Communist Czechoslovakia, but it could just as easily be today's America.

The play focuses on two artificial languages, Ptydepe and Chorukor, at the heart of the play's satire. Ptydepe was constructed along strictly scientific lines, without the ambiguities of natural languages. To avoid the possibilities for confusion that arise with homonyms, Ptydepe was created according to the postulate that all words must be formed from the least probable combinations of letters. Specifically, it makes use of the so-called "sixty percent dissimilarity" rule; which states that any Ptydepe word must differ by at least sixty percent of its letters from any other word consisting of the same number of letters. Length of words, like everything else in Ptydepe, is determined scientifically. The vocabulary of Ptydepe uses entropy encoding: shorter words have more common meanings. Therefore, the shortest word in Ptydepe, *gh*, corresponds to what is believed to be the most general term in natural language, *whatever*. (The longest word in Ptydepe, which contains 319 letters, is a name for a nonexistent member of the genus *Apus*). Havel's younger brother, computer scientist Ivan M. Havel, helped in its formulation. [Thanks to Single Carrot Theater and Wikipedia]

This performance is approximately 2 hours and 15 minutes, including a 15 minute intermission. Members of the Johns Hopkins CLSP faculty loved this play when they saw it recently and wanted to share a final performance of this extremely entertaining production with you.

6

Eighteenth Conference on Computational Natural Language Learning

Message from the Program Committee Co-Chairs

The Conference on Computational Natural Language Learning (CoNLL) is an annual meeting organized by SIGNLL, the ACL special interest group on natural language learning. CoNLL-2014 is the eighteenth meeting in this series and will be held in Baltimore, Maryland, USA on June 26-27, 2014, in conjunction with ACL-2014.

In the main session of CoNLL, we invited papers on all aspects of computational approaches to natural language learning, as well as on machine reading, the special topic of interest this year. We received 90 submissions, of which 7 were eventually withdrawn, and accepted 20 of the remaining 83 papers. The final version of each accepted paper was allowed nine content pages plus two pages with only bibliographic references.

As in previous years, CoNLL-2014 hosts a high-profile NLP shared task, Grammatical Error Correction, a continuation of the shared task in 2013. Papers of the shared task are collected in a companion volume of the CoNLL-2014 proceedings.

Similar to the arrangement in last year, each accepted paper in CoNLL-2014 will be given an 18 minute oral presentation slot, as well as a poster. There will be a poster session in the afternoon of each day, consisting of papers presented orally earlier on the same day. The shared task posters will also participate in the same poster session on Day 1.

We thank all of the authors who submitted their work to CoNLL-2014, as well as the program committee for helping us select from among the many strong submissions. We are also grateful to our invited speakers, Morten Christiansen and Tom Mitchell, who graciously agreed to give talks at CoNLL. Special thanks to the best paper award committee members, Walter Daelemans, Joakim Nivre and Dan Roth, who helped choose the best paper winner. Also thanks to Xavier Carreras and Alexander Clark, for their valuable advice, to the SIGNLL information officer, Erik Tjong Kim Sang, for publicity and to Ben Verhoeven for maintaining the CoNLL Web site. We also appreci-

ate the additional help we received from the ACL program chairs, workshop chairs, and publication chairs.

Finally, many thanks to Microsoft Research and Google for sponsoring CoNLL-2014. We hope you enjoy the conference!

Roser Morante and Scott Wen-tau Yih
CoNLL 2014 Conference Chairs

CoNLL (Main conference and shared task)
Venue: 3rd Floor Dover A/B/C

Thursday, June 26, 2014

9:00–9:18 **Opening remarks**

Session 1

9:18–9:36 What's in a p-value in NLP?
Anders Søgaard, Anders Johannsen, Barbara Plank, Dirk Hovy, and Héctor Martínez Alonso

9:36–9:54 Domain-Specific Image Captioning
Rebecca Mason and Eugene Charniak

9:54–10:12 Reconstructing Native Language Typology from Foreign Language Usage
Yevgeni Berzak, Roi Reichart, and Boris Katz

10:12–10:30 Automatic Transliteration of Romanized Dialectal Arabic
Mohamed Al-Badrashiny, Ramy Eskander, Nizar Habash, and Owen Rambow

10:30–11:00 **Coffee break**

Session 2: Shared Task

11:00–11:30 The CoNLL-2014 Shared Task on Grammatical Error Correction
Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant

11:30–11:50 Grammatical error correction using hybrid systems and type filtering
Mariano Felice, Zheng Yuan, Øistein E. Andersen, Helen Yannakoudakis, and Ekaterina Kochmar

11:50–12:10 The AMU System in the CoNLL-2014 Shared Task: Grammatical Error Correction by Data-Intensive and Feature-Rich Statistical Machine Translation
Marcin Junczys-Dowmunt and Roman Grundkiewicz

12:10–12:30 The Illinois-Columbia System in the CoNLL-2014 Shared Task
Alla Rozovskaya, Kai-Wei Chang, Mark Sammons, Dan Roth, and Nizar Habash

12:30–2:00 **Lunch break**

Session 3

2:00–2:18 Learning to Rank Answer Candidates for Automatic Resolution of Crossword Puzzles
Gianni Barlacchi, Massimo Nicosia, and Alessandro Moschitti

2:18–2:36 Inducing Neural Models of Script Knowledge
Ashutosh Modi and Ivan Titov

2:36–2:54 Grounding Language with Points and Paths in Continuous Spaces
Jacob Andreas and Dan Klein

2:54–3:12 Looking for Hyponyms in Vector Space
Marek Rei and Ted Briscoe

3:12–3:30 Lexicon Infused Phrase Embeddings for Named Entity Resolution
Alexandre Passos, Vineet Kumar, and Andrew McCallum

3:30–5:00 **Poster session 1**

- RACAI GEC – A hybrid approach to Grammatical Error Correction
  Tiberiu Boroş, Stefan Daniel Dumitrescu, Adrian Zafiu, Dan Tufiş, Verginica Barbu Mititelu, and Ionut Paul Vaduva
- Grammatical error correction using hybrid systems and type filtering
  Mariano Felice, Zheng Yuan, Øistein E. Andersen, Helen Yannakoudakis, and Ekaterina Kochmar
- Grammatical Error Detection Using Tagger Disagreement
  Anubhav Gupta
- CoNLL 2014 Shared Task: Grammatical Error Correction with a Syntactic N-gram Language Model from a Big Corpora
  S. David Hdez. and Hiram Calvo
- The AMU System in the CoNLL-2014 Shared Task: Grammatical Error Correction by Data-Intensive and Feature-Rich Statistical Machine Translation
  Marcin Junczys-Dowmunt and Roman Grundkiewicz
- Tuning a Grammar Correction System for Increased Precision
  Anoop Kunchukuttan, Sriram Chaudhury, and Pushpak Bhattacharyya
- POSTECH Grammatical Error Correction System in the CoNLL-2014 Shared Task
  Kyusong Lee and Gary Geunbae Lee
- The Illinois-Columbia System in the CoNLL-2014 Shared Task
  Alla Rozovskaya, Kai-Wei Chang, Mark Sammons, Dan Roth, and Nizar Habash
- Grammatical Error Detection and Correction using a Single Maximum Entropy Model
  Peilu Wang, Zhongye Jia, and Hai Zhao
- Factored Statistical Machine Translation for Grammatical Error Correction
  Yiming Wang, Longyue Wang, Xiaodong Zeng, Derek F. Wong, Lidia S. Chao, and Yi Lu
- NTHU at the CoNLL-2014 Shared Task
  Jian-Cheng Wu, Tzu-Hsi Yen, Jim Chang, Guan-Cheng Huang, Jimmy Chang, Hsiang-Ling Hsu, Yu-Wei Chang, and Jason S. Chang
- A Unified Framework for Grammar Error Correction
  Longkai Zhang and Houfeng Wang

5:00–6:00 **Keynote 1: Morten H. Christiansen**

Friday, June 27, 2014

8:35–9:35 **Keynote 2: Tom Mitchell**

Session 4

9:35–9:53 Focused Entailment Graphs for Open IE Propositions
Omer Levy, Ido Dagan, and Jacob Goldberger

9:53–10:11 Improved Pattern Learning for Bootstrapped Entity Extraction
Sonal Gupta and Christopher Manning

10:11–10:30 Towards Temporal Scoping of Relational Facts based on Wikipedia Data
Avirup Sil and Silviu-Petru Cucerzan

10:30–11:00 **Coffee break**

Session 5

11:00–11:18 Distributed Word Representation Learning for Cross-Lingual Dependency Parsing
Min Xiao and Yuhong Guo

11:18–11:36 Treebank Translation for Cross-Lingual Parser Induction
Jörg Tiedemann, Željko Agić, and Joakim Nivre

11:36–11:54 Weakly-Supervised Bayesian Learning of a CCG Supertagger
Dan Garrette, Chris Dyer, Jason Baldridge, and Noah A. Smith

11:54–12:12 Factored Markov Translation with Robust Modeling
Yang Feng, Trevor Cohn, and Xinkai Du

12:12–12:30 Hallucinating Phrase Translations for Low Resource MT
Ann Irvine and Chris Callison-Burch

12:30–2:00 **Lunch break**

Session 6

2:00–2:20 Linguistic Regularities in Sparse and Explicit Word Representations
Omer Levy and Yoav Goldberg

2:20–2:40 Probabilistic Modeling of Joint-context in Distributional Similarity
Oren Melamud, Ido Dagan, Jacob Goldberger, Idan Szpektor, and Deniz Yuret

2:40–3:00 A Rudimentary Lexicon and Semantics Help Bootstrap Phoneme Acquisition
Abdellah Fourtassi and Emmanuel Dupoux

3:00–3:05 **Best Paper Award announcement**

3:05–3:30 **Business meeting**

3:30–5:00 **Poster session 2**

7

Workshops

Workshop 1:
Ninth Workshop on Statistical Machine Translation (WMT14)

Organizers: *Philipp Koehn, Ondrej Bojar, Christian Buck, Christian Federmann,*
Barry Haddow, Christof Monz, Matt Post, Hervé Saint-Amand, Radu Soricut, and Lucia Specia

Venue: Salon V

Thursday, June 26, 2014

9:00–9:10 **Opening Remarks**

Session 1: Shared Translation Tasks

9:10–9:30 Efficient Elicitation of Annotations for Human Evaluation of Machine Translation
Keisuke Sakaguchi, Matt Post, and Benjamin Van Durme

9:30–10:00 Findings of the 2014 Workshop on Statistical Machine Translation
Ondrej Bojar, Christian Buck, Christian Federmann, Barry Haddow,
Philipp Koehn, Johannes Leveling, Christof Monz, Pavel Pecina,
Matt Post, Hervé Saint-Amand, Radu Soricut, Lucia Specia, and
Aleš Tamchyna

10:00–10:30 **Panel Discussion**

10:30–11:00 **Coffee**

Session 2: Poster Session

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

- Parallel FDA5 for Fast Deployment of Accurate Statistical Machine Translation Systems
  Ergun Bicici, Qun Liu, and Andy Way
- Yandex School of Data Analysis Russian-English Machine Translation System for WMT14
  Alexey Borisov and Irina Galinskaya
- CimS – The CIS and IMS joint submission to WMT 2014 translating from English into German
  Fabienne Cap, Marion Weller, Anita Ramm, and Alexander Fraser
- English-to-Hindi system description for WMT 2014: Deep Source-Context Features for Moses
  Marta R. Costa-jussà, Parth Gupta, Paolo Rosso, and Rafael E. Banchs
- The KIT-LIMSI Translation System for WMT 2014
  Quoc Khanh Do, Teresa Herrmann, Jan Niehues, Alexander Allauzen,
  François Yvon, and Alex Waibel
- The IIT Bombay Hindi-English Translation System at WMT 2014
  Piyush Dungarwal, Rajen Chatterjee, Abhijit Mishra,
  Anoop Kunchukuttan, Ritesh Shah, and Pushpak Bhattacharyya
- Edinburgh's Phrase-based Machine Translation Systems for WMT-14
  Nadir Durrani, Barry Haddow, Philipp Koehn, and Kenneth Heafield

- EU-BRIDGE MT: Combined Machine Translation
  Markus Freitag, Stephan Peitz, Joern Wuebker, Hermann Ney,
  Matthias Huck, Rico Sennrich, Nadir Durrani, Maria Nadejde,
  Philip Williams, Philipp Koehn, Teresa Herrmann, Eunah Cho, and
  Alex Waibel
- Phrasal: A Toolkit for New Directions in Statistical Machine Translation
  Spence Green, Daniel Cer, and Christopher Manning
- Anaphora Models and Reordering for Phrase-Based SMT
  Christian Hardmeier, Sara Stymne, Jörg Tiedemann, Aaron Smith, and
  Joakim Nivre
- The Karlsruhe Institute of Technology Translation Systems for the WMT
  2014
  Teresa Herrmann, Mohammed Mediani, Eunah Cho, Thanh-Le Ha,
  Jan Niehues, Isabel Slawik, Yuqi Zhang, and Alex Waibel
- The DCU-ICTCAS MT system at WMT 2014 on German-English
  Translation Task
  Liangyou Li, Xiaofeng Wu, Santiago Cortes Vaillo, Jun Xie, Andy Way, and
  Qun Liu
- The CMU Machine Translation Systems at WMT 2014
  Austin Matthews, Waleed Ammar, Archna Bhatia, Weston Feely,
  Greg Hanneman, Eva Schlinger, Swabha Swayamdipta, Yulia Tsvetkov,
  Alon Lavie, and Chris Dyer
- Stanford University's Submissions to the WMT 2014 Translation Task
  Julia Neidert, Sebastian Schuster, Spence Green, Kenneth Heafield, and
  Christopher Manning
- The RWTH Aachen German-English Machine Translation System for
  WMT 2014
  Stephan Peitz, Joern Wuebker, Markus Freitag, and Hermann Ney
- Large-scale Exact Decoding: The IMS-TTT submission to WMT14
  Daniel Quernheim and Fabienne Cap
- Abu-MaTran at WMT 2014 Translation Task: Two-step Data Selection
  and RBMT-Style Synthetic Rules
  Raphael Rubino, Antonio Toral, Víctor M. Sánchez-Cartagena,
  Jorge Ferrández-Tordera, Sergio Ortiz Rojas, Gema Ramírez-Sánchez,
  Felipe Sánchez-Martínez, and Andy Way
- The UA-Prompsit hybrid machine translation system for the 2014
  Workshop on Statistical Machine Translation
  Víctor M. Sánchez-Cartagena, Juan Antonio Pérez-Ortiz, and
  Felipe Sánchez-Martínez
- Machine Translation and Monolingual Postediting: The AFRL WMT-14
  System
  Lane Schwartz, Timothy Anderson, Jeremy Gwinnup, and Katherine Young
- CUNI in WMT14: Chimera Still Awaits Bellerophon
  Aleš Tamchyna, Martin Popel, Rudolf Rosa, and Ondrej Bojar
- Manawi: Using Multi-Word Expressions and Named Entities to Improve
  Machine Translation
  Liling Tan and Santanu Pal
- Edinburgh's Syntax-Based Systems at WMT 2014
  Philip Williams, Rico Sennrich, Maria Nadejde, Matthias Huck,
  Eva Hasler, and Philipp Koehn
- DCU-Lingo24 Participation in WMT 2014 Hindi-English Translation task
  Xiaofeng Wu, Rejwanul Haque, Tsuyoshi Okita, Piyush Arora, Andy Way,
  and Qun Liu

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

- Machine Translation of Medical Texts in the Khresmoi Project
  *Ondřej Dušek, Jan Hajič, Jaroslava Hlaváčová, Michal Novák,
  Pavel Pecina, Rudolf Rosa, Aleš Tamchyna, Zdeňka Urešová, and
  Daniel Zeman*
- Postech's System Description for Medical Text Translation Task
  Jianri Li, Se-Jong Kim, Hwidong Na, and Jong-Hyeok Lee
- Domain Adaptation for Medical Text Translation using Web Resources
  Yi Lu, Longyue Wang, Derek F. Wong, Lidia S. Chao, and Yiming Wang
- DCU Terminology Translation System for Medical Query Subtask at
  WMT14
  Tsuyoshi Okita, Ali Vahid, Andy Way, and Qun Liu
- LIMSI @ WMT'14 Medical Translation Task
  *Nicolas Pécheux, Li Gong, Quoc Khanh Do, Benjamin Marie,
  Yulia Ivanishcheva, Alexander Allauzen, Thomas Lavergne, Jan Niehues,
  Aurélien Max, and François Yvon*
- Combining Domain Adaptation Approaches for Medical Text Translation
  *Longyue Wang, Yi Lu, Derek F. Wong, Lidia S. Chao, Yiming Wang, and
  Francisco Oliveira*
- Experiments in Medical Translation Shared Task at WMT 2014
  Jian Zhang

12:30–2:00 **Lunch**

Session 3: Invited Talk

2:00–3:30 ***Machine Translation in Academia and in the Commercial World – a Contrastive Perspective.* Alon Lavie, Research Professor – Carnegie Mellon University, Co-founder, President and CTO – Safaba Translation Solutions, Inc.**

3:30–4:00 **Coffee**

Session 4: Evaluation

4:00–4:20 Randomized Significance Tests in Machine Translation
Yvette Graham, Nitika Mathur, and Timothy Baldwin

4:20–4:40 Estimating Word Alignment Quality for SMT Reordering Tasks
Sara Stymne, Jörg Tiedemann, and Joakim Nivre

4:40–5:00 Dependency-based Automatic Enumeration of Semantically Equivalent
Word Orders for Evaluating Japanese Translations
Hideki Isozaki, Natsume Kouchi, and Tsutomu Hirao

Friday, June 27, 2014

Session 5: Shared Evaluation Metrics and Quality Estimation Tasks

9:00–9:30 **Quality Estimation Shared Task**

9:30–9:50 Results of the WMT14 Metrics Shared Task
Matous Machacek and Ondrej Bojar

9:50–10:30 **Panel**

10:30–11:00 **Coffee**

Session 6: Poster Session

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

- Efforts on Machine Learning over Human-mediated Translation Edit Rate
  Eleftherios Avramidis
- SHEF-Lite 2.0: Sparse Multi-task Gaussian Processes for Translation Quality Estimation
  Daniel Beck, Kashif Shah, and Lucia Specia
- Referential Translation Machines for Predicting Translation Quality
  Ergun Bicici and Andy Way
- FBK-UPV-UEdin participation in the WMT14 Quality Estimation shared-task
  José Guilherme Camargo de Souza, Jesús González-Rubio, Christian Buck, Marco Turchi, and Matteo Negri
- Target-Centric Features for Translation Quality Estimation
  Chris Hokamp, Iacer Calixto, Joachim Wagner, and Jian Zhang
- LIG System for Word Level QE task at WMT14
  Ngoc Quang Luong, Laurent Besacier, and Benjamin Lecouteux
- Exploring Consensus in Machine Translation for Quality Estimation
  Carolina Scarton and Lucia Specia
- LIMSI Submission for WMT'14 QE Task
  Guillaume Wisniewski, Nicolas Pécheux, Alexander Allauzen, and François Yvon

11:00–12:30 **Shared Task: Evaluation Metrics**

- Parmesan: Meteor without Paraphrases with Paraphrased References
  Petra Barancikova
- A Systematic Comparison of Smoothing Techniques for Sentence-Level BLEU
  Boxing Chen and Colin Cherry
- VERTa participation in the WMT14 Metrics Task
  Elisabet Comelles and Jordi Atserias
- Meteor Universal: Language Specific Translation Evaluation for Any Target Language
  Michael Denkowski and Alon Lavie
- Application of Prize based on Sentence Length in Chunk-based Automatic Evaluation of Machine Translation
  Hiroshi Echizen'ya, Kenji Araki, and Eduard Hovy
- LAYERED: Metric for Machine Translation Evaluation
  Shubham Gautam and Pushpak Bhattacharyya
- IPA and STOUT: Leveraging Linguistic and Source-based Features for Machine Translation Evaluation
  Meritxell Gonzàlez, Alberto Barrón-Cedeño, and Lluís Màrquez
- DiscoTK: Using Discourse Structure for Machine Translation Evaluation
  Shafiq Joty, Francisco Guzmán, Lluís Màrquez, and Preslav Nakov
- Tolerant BLEU: a Submission to the WMT14 Metrics Task
  Jindřich Libovický and Pavel Pecina
- BEER: BEtter Evaluation as Ranking
  Milos Stanojevic and Khalil Sima'an
- RED, The DCU-CASICT Submission of Metrics Tasks
  Xiaofeng Wu, Hui Yu, and Qun Liu

12:30–2:00 **Lunch**

Session 7: Data and Adaptation

2:00–2:20 Crowdsourcing High-Quality Parallel Data Extraction from Twitter
Wang Ling, Luis Marujo, Chris Dyer, Alan W Black, and Isabel Trancoso

2:20–2:40 Using Comparable Corpora to Adapt MT Models to New Domains
Ann Irvine and Chris Callison-Burch

Workshop 2:
BioNLP

Organizers: *Kevin Cohen, Dina Demner-Fushman, Sophia Ananiadou, John Pestian,*
Jun-ichi Tsujii, Thamar Solorio, and Yang Liu

Venue: Salon III–IV

Thursday, June 26, 2014

9:00–9:10 **Opening remarks**

Session 1: Processing biomedical publications

9:10–9:30 Natural Language Processing Methods for Enhancing Geographic
Metadata for Phylogeography of Zoonotic Viruses
Tasnia Tahsin, Robert Rivera, Rachel Beard, Rob Lauder,
Davy Weissenbacher, Matthew Scotch, Garrick Wallstrom, and
Graciela Gonzalez

9:30–9:50 Temporal Expression Recognition for Cell Cycle Phase Concepts in
Biomedical Literature
Negacy Hailu, Natalya Panteleyeva, and Kevin Cohen

9:50–10:10 Classifying Negative Findings in Biomedical Publications
Bei Yu

10:10–10:30 Automated Disease Normalization with Low Rank Approximations
Robert Leaman and Zhiyong Lu

10:30–11:00 **Coffee Break**

Keynote by Junichi Tsujii

11:00–11:50 **BioNLP as the Pioneering field of linking text, knowledge and data**

Session 2: Processing consumer language

11:50–12:10 Decomposing Consumer Health Questions
Kirk Roberts, Halil Kilicoglu, Marcelo Fiszman, and
Dina Demner-Fushman

12:10–12:30 Detecting Health Related Discussions in Everyday Telephone
Conversations for Studying Medical Events in the Lives of Older Adults
Golnar Sheikhshab, Izhak Shafran, and Jeffrey Kaye

12:30–2:00 **Lunch**

Session 3: Processing clinical text and gray literature

2:00–2:20 Coreference Resolution for Structured Drug Product Labels
Halil Kilicoglu and Dina Demner-Fushman

2:20–2:40 Generating Patient Problem Lists from the ShARe Corpus using
SNOMED CT/SNOMED CT CORE Problem List
Danielle Mowery, Mindy Ross, Sumithra Velupillai, Stephane Meystre,
Janyce Wiebe, and Wendy Chapman

2:40–3:00 A System for Predicting ICD-10-PCS Codes from Electronic Health
Records
Michael Subotin and Anthony Davis

3:00–3:20 Structuring Operative Notes using Active Learning
Kirk Roberts, Sanda Harabagiu, and Michael Skinner

3:30–4:00 **Afternoon Break**

4:00–4:20 Chunking Clinical Text Containing Non-Canonical Language
Aleksandar Savkov, John Carroll, and Jackie Cassell

4:20–4:40 Decision Style in a Clinical Reasoning Corpus
*Limor Hochberg, Cecilia Ovesdotter Alm, Esa M. Rantanen,
Caroline M. DeLong, and Anne Haake*

(4:40–5:30) **Poster session**

- Temporal Expressions in Swedish Medical Text – A Pilot Study
  Sumithra Velupillai
- A repository of semantic types in the MIMIC II database clinical notes
  Richard Osborne, Alan Aronson, and Kevin Cohen
- Extracting drug indications and adverse drug reactions from Spanish
  health social media
  *Isabel Segura-Bedmar, Santiago de la Peña González, and
  Paloma Martínez*
- Symptom extraction issue
  Laure Martin, Delphine Battistelli, and Thierry Charnois
- Seeking Informativeness in Literature Based Discovery
  Judita Preiss
- Towards Gene Recognition from Rare and Ambiguous Abbreviations
  using a Filtering Approach
  Matthias Hartung, Roman Klinger, Matthias Zwick, and Philipp Cimiano
- FFTM: A Fuzzy Feature Transformation Method for Medical Documents
  Amir Karami and Aryya Gangopadhyay

Friday, June 27, 2014

Session 1: NLP approaches for assessment of clinical conditions

9:00–9:40 Using statistical parsing to detect agrammatic aphasia
*Kathleen C. Fraser, Graeme Hirst, Jed A. Meltzer, Jennifer E. Mack, and
Cynthia K. Thompson*

**Panel: Life cycles of BioCreative, BioNLP-ST, i2b2, TREC Medical
tracks, and ShARe /CLEF/ SemEval**

9:40–10:05 **BioCreative by Lynette Hirschman and John Wilbur**

10:05–10:30 **BioNLP-ST by Sophia Ananiadou and Junichi Tsujii**

10:30–11:00 **Coffee Break**

11:00–11:25 **TREC Medical tracks by Ellen Voorhees**

11:25–11:50 **i2b2 by Ozlem Uzuner**

11:50–12:10 **ShARe/CLEF/SemEval by Danielle Mowery, Sumithra Velupillai and
Sameer Pradhan**

12:10–12:30 **Discussion**

12:30–2:00 **Lunch**

Tutorials

2:00–3:30 **UMLS in biomedical text processing by Olivier Bodenreider**

3:30–4:00 **Afternoon Break**

4:00–5:30 **Using MetaMap by Alan R. Aronson**

Workshop 3:
9th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)

Organizers: *Joel Tetreault, Jill Burstein, and Claudia Leacock*

Venue: Salon VIII

Thursday, June 26, 2014

8:45–9:00 **Load Presentations**

9:00–9:15 **Opening Remarks**

9:15–9:40 Automated Measures of Specific Vocabulary Knowledge from Constructed Responses ('Use These Words to Write a Sentence Based on this Picture')
Swapna Somasundaran and Martin Chodorow

9:40–10:05 Automatic Assessment of the Speech of Young English Learners
Jian Cheng, Yuan Zhao D'Antilio, Xin Chen, and Jared Bernstein

10:05–10:25 Automatic detection of plagiarized spoken responses
Keelan Evanini and Xinhao Wang

10:30–11:00 **Break**

11:00–11:20 Understanding MOOC Discussion Forums using Seeded LDA
Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daume, and Lise Getoor

11:20–12:30 **Invited Speaker: Norbert Elliot**

12:20–2:00 **Lunch**

2:00–3:30 **Posters and Demos**

- Translation Class Instruction as Collaboration in the Act of Translation
  Lars Ahrenberg and Ljuba Tarvi
- The pragmatics of margin comments: An empirical study
  Debora Field, Stephen Pulman, and Denise Whitelock
- Surprisal as a Predictor of Essay Quality
  Gaurav Kharkwal and Smaranda Muresan
- Towards Domain-Independent Assessment of Elementary Students' Science Competency using Soft Cardinality
  Samuel Leeman-Munk, Angela Shelton, Eric Wiebe, and James Lester
- Automatic evaluation of spoken summaries: the case of language assessment
  Anastassia Loukina, Klaus Zechner, and Lei Chen
- An Explicit Feedback System for Preposition Errors based on Wikipedia Revisions
  Nitin Madnani and Aoife Cahill
- Syllable and language model based features for detecting non-scorable tests in spoken language proficiency assessment applications
  Angeliki Metallinou and Jian Cheng
- Improving Peer Feedback Prediction: The Sentence Level is Right
  Huy Nguyen and Diane Litman

- ArCADE: An Arabic Corpus of Auditory Dictation Errors
  C. Anton Rytting, Paul Rodrigues, Tim Buckwalter, Valerie Novak, Aric Bills, Noah H. Silbert, and Mohini Madgavkar
- Similarity-Based Non-Scorable Response Detection for Automated Speech Scoring
  Su-Youn Yoon and Shasha Xie
- Natural Language Generation with Vocabulary Constraints
  Ben Swanson, Elif Yamangil, and Eugene Charniak
- Automated scoring of speaking items in an assessment for teachers of English as a Foreign Language
  Klaus Zechner, Keelan Evanini, Su-Youn Yoon, Lawrence Davis, Xinhao Wang, Lei Chen, Chong Min Lee, and Chee Wee Leong
- Automatic Generation of Challenging Distractors Using Context-Sensitive Inference Rules
  Torsten Zesch and Oren Melamud
- Sentence-level Rewriting Detection
  Fan Zhang and Diane Litman

3:30–4:00 **Break**

4:00–4:25 Exploiting Morphological, Grammatical, and Semantic Correlates for Improved Text Difficulty Assessment
Elizabeth Salesky and Wade Shen

4:25–4:50 Assessing the Readability of Sentences: Which Corpora and Features?
Felice Dell'Orletta, Martijn Wieling, Giulia Venturi, Andrea Cimino, and Simonetta Montemagni

4:50–5:15 Rule-based and machine learning approaches for second language sentence-level readability
Ildikó Pilán, Elena Volodina, and Richard Johansson

5:15–5:30 **Closing Remarks**

Workshop 4:
Speech and Language Processing for Assistive Technologies (SLPAT 2014)

Organizers: *Jan Alexandersson, Dimitra Anastasiou, Cui Jian, Ani Nenkova, Rupal Patel, Frank Rudzicz, Annalu Waller, and Desislava Zhekova*

Venue: Salon I

Thursday, June 26, 2014

9:00–9:15 **Welcome & Opening Remarks**

Session 1

9:14–9:45 Standing on the shoulders of giants: attacking the meta-problems of technical AAC research
Joseph Reddington

9:45–10:15 Graphical Modification of Text. An Approach To Dyslexic Users.
Tereza Pařilová

10:15–10:30 **Break**

Session 2

10:30–11:00 **Special Talk by K.I. Boser, M. Goodwin, and S.C. Wayland:** *Technology Tools For Students With Autism: Innovations That Enhance Independence and Learning*

11:00–11:30 Dialogue Strategy Learning in Healthcare: A Systematic Approach for Learning Dialogue Models from Data
Hamidreza Chinaei and Brahim Chaib-draa

11:30–12:00 Speech recognition in Alzheimer's disease with personal assistive robots
Frank Rudzicz, Rosalie Wang, Momotaz Begum, and Alex Mihailidis

12:00–2:00 **Lunch**

Session 3

2:00–2:30 Individuality-preserving Voice Conversion for Articulation Disorders Using Dictionary Selective Non-negative Matrix Factorization
Ryo Aihara, Tetsuya Takiguchi, and Yasuo Ariki

2:30–3:00 Preliminary Test of a Real-Time, Interactive Silent Speech Interface Based on Electromagnetic Articulograph
Jun Wang, Ashok Samal, and Jordan Green

3:00–4:00 **Invited Talk by Frank Rudzicz:** *Dysarthria as a noisy channel in speech production*

Workshop 5:
Cognitive Modeling and Computational Linguistics (CMCL)

Organizers: *Vera Demberg and Tim O'Donnell*

Venue: Salon VII

Thursday, June 26, 2014

8:55–9:00 **Opening**

9:00–10:00 **Invited talk by Naomi Feldman**

10:00–10:30 Computationally Rational Saccadic Control: An Explanation of Spillover Effects Based on Sampling from Noisy Perception and Memory
Michael Shvartsman, Richard Lewis, and Satinder Singh

10:30–11:00 **Coffee break**

11:00–11:30 Investigating the role of entropy in sentence processing
Tal Linzen and Florian T. Jaeger

11:30–12:00 Sentence Processing in a Vectorial Model of Working Memory
William Schuler

12:00–12:30 Evaluating Evaluation Metrics for Minimalist Parsing
Thomas Graf and Bradley Marcinek

12:30–2:00 **Lunch break**

2:00–2:30 Learning Verb Classes in an Incremental Model
Libby Barak, Afsaneh Fazly, and Suzanne Stevenson

2:30–3:00 A Usage-Based Model of Early Grammatical Development
Barend Beekhuizen, Rens Bod, Afsaneh Fazly, Suzanne Stevenson, and Arie Verhagen

3:00–3:30 A Model to Qualify the Linguistic Adaptation Phenomenon in Online Conversation Threads: Analyzing Priming Effect in Online Health Community
Yafei Wang, David Reitter, and John Yen

3:30–4:00 **Coffee break**

4:00–4:30 Quantifying the role of discourse topicality in speakers' choices of referring expressions
Naho Orita, Naomi H. Feldman, Jordan Boyd-Graber, and Eliana Vornov

4:30–5:30 **Invited talk by Ted Gibson**

5:30–5:45 **Closing Remarks**

Workshop 6:
Argumentation Mining

Organizers: *Kevin Ashley, Nancy Green, Diane Litman, Chris Reed, and Vern Walker*

Venue: Salon IX

Thursday, June 26, 2014

Session 1: Papers

8:30–9:00 Annotating Patterns of Reasoning about Medical Theories of Causation in Vaccine Cases: Toward a Type System for Arguments
Vern Walker, Karina Vazirova, and Cass Sanford

9:00–9:30 Towards Creation of a Corpus for Argumentation Mining the Biomedical Genetics Research Literature
Nancy Green

9:30–10:00 An automated method to build a corpus of rhetorically-classified sentences in biomedical texts
Hospice Houngbo and Robert Mercer

10:00–10:30 Ontology-Based Argument Mining and Automatic Essay Scoring
Nathan Ong, Diane Litman, and Alexandra Brusilovsky

10:30–11:00 **Coffee**

Session 2: Papers

10:30–11:00 Identifying Appropriate Support for Propositions in Online User Comments
Joonsuk Park and Claire Cardie

11:00–11:30 Analyzing Argumentative Discourse Units in Online Interactions
Debanjan Ghosh, Smaranda Muresan, Nina Wacholder, Mark Aakhus, and Matthew Mitsui

11:30–12:00 Back up your Stance: Recognizing Arguments in Online Discussions
Filip Boltužić and Jan Šnajder

12:00–12:30 Automated argumentation mining to the rescue? Envisioning argumentation and decision-making support for debates in open online collaboration communities
Jodi Schneider

12:30–2:00 **Lunch**

Session 3: Papers

2:00–2:20 A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics
Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim

2:20–2:40 Applying Argumentation Schemes for Essay Scoring
Yi Song, Michael Heilman, Beata Beigman Klebanov, and Paul Deane

2:40–3:00 Mining Arguments From 19th Century Philosophical Texts Using Topic Based Modelling
John Lawrence, Chris Reed, Colin Allen, Simon McAlister, and Andrew Ravenscroft

3:00–3:20 Towards segment-based recognition of argumentation structure in short texts
Andreas Peldszus

3:30–4:00 **Coffee**

Session 4: Posters

4:00–5:00 **Poster session**

- Titles That Announce Argumentative Claims in Biomedical Research Articles
  Heather Graves, Roger Graves, Robert Mercer, and Mahzereen Akter
- Extracting Higher Order Relations From Biomedical Text
  Syeed Ibn Faiz and Robert Mercer
- Survey in sentiment, polarity and function analysis of citation
  Myriam Hernández A. and José M. Gómez
- Indicators of Argument-conclusion Relationships. An Approach for Argumentation Mining in German Discourses
  Bianka Trevisan, Eva Dickmeis, Eva-Maria Jakobs, and Thomas Niehr
- Extracting Imperatives from Wikipedia Article for Deletion Discussions
  Fiona Mao, Robert Mercer, and Lu Xiao
- Requirement Mining in Technical Documents
  Juyeon Kang and Patrick Saint-Dizier

Workshop 7:

ComputEL: The use of computational methods in the study of endangered languages

Organizers: *Jeff Good, Julia Hirschberg, and Owen Rambow*

Venue: Heron (4th floor)

Thursday, June 26, 2014

9:00–9:10	**Introduction**
	Paper Session 1: Computational Tools for Endangered Languages Research
9:10–9:30	Aikuma: A Mobile App for Collaborative Language Documentation *Steven Bird, Florian R. Hanke, Oliver Adams, and Haejoong Lee*
9:30–9:50	Documenting Endangered Languages with the WordsEye Linguistics Tool *Morgan Ulinski, Anusha Balakrishnan, Daniel Bauer, Bob Coyne, Julia Hirschberg, and Owen Rambow*
9:50–10:10	Small Languages, Big Data: Multilingual Computational Tools and Techniques for the Lexicography of Endangered Languages *Martin Benjamin and Paula Radetzky*
10:10–10:30	LingSync & the Online Linguistic Database: New Models for the Collection and Management of Data for Language Communities, Linguists and Language Learners *Joel Dunham, Gina Cook, and Joshua Horner*
10:30–11:00	**Coffee Break**
	Paper Session 2: Applying Computational Methods to Endangered Languages
11:00–11:30	Modeling the Noun Morphology of Plains Cree *Conor Snoek, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen, and Trond Trosterud*
11:30–12:00	Learning Grammar Specifications from IGT: A Case Study of Chintang *Emily M. Bender, Joshua Crowgey, Michael Wayne Goodman, and Fei Xia*
12:00–12:30	Creating Lexical Resources for Endangered Languages *Khang Nhut Lam, Feras Al Tarouti, and Jugal Kalita*
12:30–2:00	**Lunch**
2:00–3:00	**Posters and Demonstrations of Tools Presented in Paper Session 1**
	• Estimating Native Vocabulary Size in an Endangered Language *Timofey Arkhangelskiy*
	• InterlinguaPlus Machine Translation Approach for Local Languages: Ekegusii & Swahili *Edward Ombui, Peter Wagacha, and Wanjiku Ng'ang'a*
	• Building and Evaluating Somali Language Corpora *Abdillahi Nimaan*
	Paper Session 3: Infrastructure and Community Development for Computational Research on Endangered Languages

3:00–3:30 SeedLing: Building and Using a Seed Corpus for the Human Language Project
Guy Emerson, Liling Tan, Susanne Fertmann, Alexis Palmer, and Michaela Regneri

3:30–4:00 **Coffee Break**

4:00–4:20 Short-Term Projects, Long-Term Benefits: Four Student NLP Projects for Low-Resource Languages
Alexis Palmer and Michaela Regneri

4:20–4:50 Data Warehouse, Bronze, Gold, STEC, Software
Doug Cooper

4:50–5:20 Time to Change the "D" in "DEL"
Stephen Beale

5:20–5:30 **Concluding Remarks**

Workshop 8:
The Second Workshop on Metaphor in NLP

Organizers: *Beata Beigman Klebanov, Ekaterina Shutova, and Patricia Lichtenstein*

Venue: Galena (4th floor)

Thursday, June 26, 2014

9:00–9:05 **Opening Remarks**

9:05–10:00 **Invited talk: Brad Pasanek "Giving Back the Image of the Mind: Computational Approaches to 'Propriety' in Eighteenth-Century British Literature"**

10:00–10:30 Conversational Metaphors in Use: Exploring the Contrast between Technical and Everyday Notions of Metaphor
Hyeju Jang, Mario Piergallini, Miaomiao Wen, and Carolyn Rose

10:30–11:00 **Coffee Break**

11:00–11:30 Different Texts, Same Metaphors: Unigrams and Beyond
Beata Beigman Klebanov, Ben Leong, Michael Heilman, and Michael Flor

11:30–12:00 Metaphor Detection through Term Relevance
Marc Schulder and Eduard Hovy

12:00–12:30 Multi-dimensional abstractness in cross-domain mappings
Jonathan Dunn

12:30–2:00 **Lunch**

2:00–2:30 Abductive Inference for Interpretation of Metaphors
*Ekaterina Ovchinnikova, Ross Israel, Suzanne Wertheim,
Vladimir Zaytsev, Niloofar Montazeri, and Jerry Hobbs*

2:30–3:00 Computing Affect in Metaphors
*Tomek Strzalkowski, Samira Shaikh, Kit Cho, George Aaron Broadwell,
Laurie Feldman, Sarah Taylor, Boris Yamrom, Ting Liu, Ignacio Cases,
Yuliya Peshkova, and Kyle Elliot*

3:00–3:30 A Service-Oriented Architecture for Metaphor Processing
Tony Veale

3:30–3:45 **Coffee Break**

3:45–4:30 **Invited talk: Rebecca Resnik "Pandora's Box: Uses of metaphor in clinical psychology and computational linguistics"**

4:30–5:30 **Panel discussion: "Metaphors We Work On: Goals, Trajectories and Applications"**

Workshop 9:
Workshop on Semantic Parsing (SP14)

Organizers: *Yoav Artzi, Tom Kwiatkowski, and Jonathan Berant*

Venue: Salon VI

Thursday, June 26, 2014

9:00–9:05 **Opening Remarks**

Invited Talks

9:05–9:50 *Semantic Parsing: Past, Present, and Future* **Raymond Mooney**

9:50–10:20 *Can a Machine Translate Without Knowing What Translation Is?* **Kevin Knight**

Exceptional Submission Talks

10:20–10:30 Low-Dimensional Embeddings of Logic
Tim Rocktäschel, Matko Bošnjak, Sameer Singh, and Sebastian Riedel

10:30–11:00 **Coffee Break**

11:00–11:10 Combining Formal and Distributional Models of Temporal and Intensional Semantics
Mike Lewis and Mark Steedman

11:10–11:20 Cooking with Semantics
Jonathan Malmaud, Earl Wagner, Nancy Chang, and Kevin Murphy

11:20–12:30 **Poster Session**

- Learning a Lexicon for Broad-coverage Semantic Parsing
  James Allen
- Semantic Parsing using Distributional Semantics and Probabilistic Logic
  Islam Beltagy, Katrin Erk, and Raymond Mooney
- Large-scale CCG Induction from the Groningen Meaning Bank
  Sebastian Beschke, Yang Liu, and Wolfgang Menzel
- Semantic Parsing for Text to 3D Scene Generation
  Angel Chang, Manolis Savva, and Christopher Manning
- Leveraging Frame Semantics and Distributional Semantics for Unsupervised Semantic Slot Induction in Spoken Dialogue Systems
  Yun-nung Chen, William Yang Wang, and Alexander Rudnicky
- Semantic Parsing for Information Extraction
  Eunsol Choi, Tom Kwiatkowski, and Luke Zettlemoyer
- Parsing and Grounding Referring Expressions in Automatically Constructed 3D Models
  Nicholas FitzGerald and Luke Zettlemoyer
- Graph-Based Algorithms for Semantic Parsing
  Jeffrey Flanigan, Samuel Thomson, David Bamman, Jesse Dodge, Manaal Faruqui, Brendan O'Connor, Nathan Schneider, Swabha Swayamdipta, Chris Dyer, and Noah A. Smith
- A Deep Architecture for Semantic Parsing
  Edward Grefenstette, Phil Blunsom, Nando de Freitas, and Karl Moritz Hermann

- Symmetry-Based Semantic Parsing
  Chloé Kiddon and Pedro Domingos
- Leveraging Heterogeneous Data Sources for Relational Semantic Parsing
  Meghana Kshirsagar, Nathan Schneider, and Chris Dyer
- Context-dependent Semantic Parsing for Time Expressions
  Kenton Lee, Yoav Artzi, Jesse Dodge, and Luke Zettlemoyer
- Combining Formal and Distributional Models of Temporal and Intensional Semantics
  Mike Lewis and Mark Steedman
- Cooking with Semantics
  Jonathan Malmaud, Earl Wagner, Nancy Chang, and Kevin Murphy
- Representing Caused Motion in Embodied Construction Grammar
  Ellen K Dodge and Miriam R.L. Petruck
- Constructing Semantic Parsing Datasets from Technical Documentation
  Kyle Richardson and Jonas Kuhn
- Notes on the MCTest Dataset for the Open-Domain Machine Comprehension of Text
  Matthew Richardson, Christopher J.C. Burges, and Erin Renshaw
- Low-Dimensional Embeddings of Logic
  Tim Rocktäschel, Matko Bošnjak, Sameer Singh, and Sebastian Riedel
- Software Requirements: A new Domain for Semantic Parsers
  Michael Roth, Themistoklis Diamantopoulos, Ewan Klein, and Andreas Symeonidis
- From Treebank Parses to Episodic Logic and Commonsense Inference
  Lenhart Schubert
- On maximum spanning DAG algorithms for semantic DAG parsing
  Natalie Schluter
- Intermediary Semantic Representation through Proposition Structures
  Gabriel Stanovsky, Jessica Ficler, Ido Dagan, and Yoav Goldberg
- Efficient Logical Inference for Semantic Processing
  Ran Tian, Yusuke Miyao, and Takuya Matsuzaki
- A New Corpus for Context-Dependent Semantic Parsing
  Andreas Vlachos and Stephen Clark
- Towards README-EVAL : Interpreting README File Instructions
  James White
- Freebase QA: Information Extraction or Semantic Parsing?
  Xuchen Yao, Jonathan Berant, and Benjamin Van Durme

12:30–2:10 **Lunch Break**

Invited Talks

2:10–2:50 ***Semantic Parsing for Cancer Panomics*** **Hoifung Poon**

2:50–3:30 ***Semantics for Semantic Parsers*** **Mark Steedman**

3:30–4:00 **Coffee Break**

4:00–4:40 ***Asking for Help Using Inverse Semantics*** **Stefanie Tellex**

4:40–5:20 ***Computing with Natural Language*** **Percy Liang**

5:20–6:00 ***Grounded Semantic Parsing*** **Luke Zettlemoyer**

6:00–6:00 **Closing**

Workshop 10:
Language Technologies and Computational Social Science

Organizers: *Cristian Danescu-Niculescu-Mizil, Jacob Eisenstein, Kathleen McKeown, and Noah A. Smith*

Venue: Salon II

Thursday, June 26, 2014

Welcome; Invited Talks, Style and Rhetoric

9:00–9:05 **Welcome**

9:05–9:35 Is It All in the Phrasing? Computational Explorations in How We Say What We Say, and Why It Matters
Lillian Lee

9:35–10:05 Creating and Destroying Party Brands
Justin Grimmer

10:05–10:30 **Discussion**

10:30–11:00 **Coffee break**

Invited Talks, Sociolinguistics and Social Media

11:00–11:30 Sociolinguistics for Computational Social Science
Sali Tagliamonte

11:30–12:00 Location and Language Use in Social Media
Ed Chi

12:00–12:30 **Discussion**

12:30–2:00 **Lunch break**

Unshared task

2:00–2:30 Overview of the 2014 NLP Unshared Task in PoliInformatics
Noah A. Smith, Claire Cardie, Anne Washington, and John Wilkerson

2:30–3:30 **Poster session**

- Context-based Natural Language Processing for GIS-based Vague Region Visualization
  Wei Chen
- Extracting Socioeconomic Patterns from the News: Modelling Text and Outlet Importance Jointly
  Vasileios Lampos, Daniel Preoţiuc-Pietro, Sina Samangooei, Douwe Gelling, and Trevor Cohn
- Fact Checking: Task definition and dataset construction
  Andreas Vlachos and Sebastian Riedel
- Finding Eyewitness Tweets During Crises
  Fred Morstatter, Nichola Lubold, Heather Pon-Barry, Jürgen Pfeffer, and Huan Liu
- Inducing Information Structures for Data-driven Text Analysis
  Andrew Salway, Samia Touileb, and Endre Tvinnereim
- Information density, Heaps' Law, and perception of factiness in news
  Miriam Boon

- Measuring the Public Accountability of New Modes of Governance
  Bruno Wueest, Gerold Schneider, and Michael Amsler
- Optimizing Features in Active Machine Learning for Complex Qualitative Content Analysis
  Jasy Suet Yan Liew, Nancy McCracken, Shichun Zhou, and Kevin Crowston
- Power of Confidence: How Poll Scores Impact Topic Dynamics in Political Debates
  Vinodkumar Prabhakaran, Ashima Arora, and Owen Rambow
- Predicting Fine-grained Social Roles with Selectional Preferences
  Charley Beller, Craig Harman, and Benjamin Van Durme
- Predicting Party Affiliations from European Parliament Debates
  Bjørn Høyland, Jean-François Godbout, Emanuele Lapponi, and Erik Velldal
- Temporal Analysis of Language through Neural Language Models
  Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov
- Using Simple NLP Tools to Trace the Globalization of the Art World
  Mohamed AlTantawy, Alix Rule, Owen Rambow, Zhongyu Wang, and Rupayan Basu

3:30–4:00 **Coffee break**

Invited Talks, Framing and Agenda-Setting

4:00–4:30 Issue Framing as a Generalizable Phenomenon
Amber Boydstun

4:30–5:00 "I Want to Talk About, Again, My Record On Energy ...": Modeling Agendas and Framing in Political Debates and Other Conversations
Philip Resnik

5:00–5:30 **Discussion**

5:30–5:45 **Closing remarks**

Workshop 11:

5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2014)

Organizers: *Alexandra Balahur, Erik van der Goot, Ralf Steinberger, and Andrés Montoyo*

Venue: Salon IX

Friday, June 27, 2014

8:30–8:35 **Opening Remarks**

8:35–9:10 **Invited talk: Dr. Saif Mohammad**

8:35–9:10 Words: Evaluative, Emotional, Colourful, Musical!
Saif Mohammad

Session 1: Cross-domain and Multilingual Sentiment Analysis

9:10–9:35 Robust Cross-Domain Sentiment Analysis for Low-Resource Languages
Jakob Elming, Barbara Plank, and Dirk Hovy

9:35–10:00 An Investigation for Implicatures in Chinese : Implicatures in Chinese and in English are similar !
Lingjia Deng and Janyce Wiebe

10:00–10:15 Inducing Domain-specific Noun Polarity Guided by Domain-independent Polarity Preferences of Adjectives
Manfred Klenner, Michael Amsler, and Nora Hollenstein

10:15–10:30 Aspect-Level Sentiment Analysis in Czech
Josef Steinberger, Tomáš Brychcín, and Michal Konkol

10:30–10:50 **Break**

10:50–11:25 **Invited talk: Dr. Myle Ott**

10:50–11:25 Linguistic Models of Deceptive Opinion Spam
Myle Ott

Session 2: Emotion, Irony and Sarcasm Classification

11:25–11:50 Semantic Role Labeling of Emotions in Tweets
Saif Mohammad, Xiaodan Zhu, and Joel Martin

11:50–12:15 An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews
Konstantin Buschmeier, Philipp Cimiano, and Roman Klinger

12:15–12:30 Modelling Sarcasm in Twitter, a Novel Approach
Francesco Barbieri, Horacio Saggion, and Francesco Ronzano

12:30–12:45 Emotive or Non-emotive: That is The Question
Michal Ptaszynski, Fumito Masui, Rafal Rzepka, and Kenji Araki

12:45–2:00 **Lunch Break**

2:00–2:30 **Demo talk: Dr. Alexandra Balahur**

2:00–2:30 Challenges in Creating a Multilingual Sentiment Analysis Application for Social Media Mining
Alexandra Balahur, Hristo Tanev, and Erik van der Goot

Session 3: Lexical Acquisition and Feature Weighting for Sentiment

Analysis

2:30–2:45 Two-Step Model for Sentiment Lexicon Extraction from Twitter Streams
Ilia Chetviorkin and Natalia Loukachevitch

2:45–3:00 Linguistically Informed Tweet Categorization for Online Reputation
Management
Gerard Lynch and Pádraig Cunningham

3:00–3:15 Credibility Adjusted Term Frequency: A Supervised Term Weighting
Scheme for Sentiment Analysis and Text Classification
Yoon Kim and Owen Zhang

3:15–3:30 Opinion Mining and Topic Categorization with Novel Term Weighting
*Tatiana Gasanova, Roman Sergienko, Shakhnaz Akhmedova,
Eugene Semenkin, and Wolfgang Minker*

3:30–4:00 **Break**

Session 4: Sentiment Analysis from Discourse and Dialogues

4:00–4:25 Sentiment classification of online political discussions: a comparison of a
word-based and dependency-based method
Hugo Lewi Hammer, Per Erik Solberg, and Lilja Øvrelid

4:25–4:50 Improving Agreement and Disagreement Identification in Online
Discussions with A Socially-Tuned Sentiment Lexicon
Lu Wang and Claire Cardie

4:50–5:05 Lexical Acquisition for Opinion Inference: A Sense-Level Lexicon of
Benefactive and Malefactive Events
Yoonjung Choi, Lingjia Deng, and Janyce Wiebe

5:05–5:20 Dive deeper: Deep Semantics for Sentiment Analysis
Nikhilkumar Jadhav and Pushpak Bhattacharyya

5:20–5:30 **Break**

Session 5: Sentiment Analysis Evaluation. Going Beyond Current Sentiment Analysis Approaches

5:30–5:55 Evaluating Sentiment Analysis Evaluation: A Case Study in Securities
Trading
Siavash Kazemian, Shunan Zhao, and Gerald Penn

5:55–6:20 Sentiment Classification on Polarity Reviews: An Empirical Study Using
Rating-based Features
Dai Quoc Nguyen, Dat Quoc Nguyen, Thanh Vu, and Son Bao Pham

6:20–6:35 Effect of Using Regression on Class Confidence Scores in Sentiment
Analysis of Twitter Data
Itir Onal, Ali Mert Ertugrul, and Ruken Cakici

6:35–6:50 A cognitive study of subjectivity extraction in sentiment annotation
Abhijit Mishra, Aditya Joshi, and Pushpak Bhattacharyya

6:50–7:05 The Use of Text Similarity and Sentiment Analysis to Examine Rationales
in the Large-Scale Online Deliberations
Wanting Mao, Lu Xiao, and Robert Mercer

7:05–7:20 A Conceptual Framework for Inferring Implicatures
Janyce Wiebe and Lingjia Deng

7:20–7:30 **Closing remarks**

Workshop 12:
Joint Workshop on Social Dynamics and Personal Attributes in Social Media

Organizers: *Alice Oh, Oren Tsur, Benjamin Van Durme, Svitlana Volkova, and David Yarowsky*

Venue: Salon VIII

Friday, June 27, 2014

8:40–8:50 **Welcome Notes**

8:50–9:50 **Invited talk by Derek Ruths**

Oral Session 1

10:05–10:25 Detecting Sociostructural Beliefs about Group Status Differences in Online Discussions
Brian Riordan, Heather Wade, and Afzal Upal

10:25–10:45 Using County Demographics to Infer Attributes of Twitter Users
Ehsan Mohammady and Aron Culotta

Oral Session 2

10:55–11:10 The Enrollment Effect: A Study of Amazon's Vine Program
Dinesh Puranam and Claire Cardie

11:10–11:30 Discourse Analysis of User Forums in an Online Weight Loss Application
Lydia Manikonda, Heather Pon-Barry, Subbarao Kambhampati, Eric Hekler, and David W. McDonald

11:30–12:30 **Panel by Hanna Wallach, Jacob Eisenstein, Cristian Danescu-Niculescu-Mizil and Jimmy Lin**

Oral Session 3

1:30–1:50 A Unified Topic-Style Model for Online Discussions
Ying Ding, Jing Jiang, and Qiming Diao

1:50–2:10 Self-disclosure topic model for Twitter conversations
JinYeong Bak, Chin-Yew Lin, and Alice Oh

2:10–2:25 Detecting and Evaluating Local Text Reuse in Social Networks
Shaobin Xu, David Smith, Abigail Mullen, and Ryan Cordell

2:35–3:35 **Invited talk**

Oral Session 4

3:45–4:00 Generating Subjective Responses to Opinionated Articles in Social Media: An Agenda-Driven Architecture and a Turing-Like Test
Tomer Cagan, Stefan L. Frank, and Reut Tsarfaty

4:00–4:15 A Semi-Automated Method of Network Text Analysis Applied to 150 Original Screenplays
Starling Hunter

Poster Session

4:30–5:30 Power of Confidence: How Poll Scores Impact Topic Dynamics in Political Debates
Vinodkumar Prabhakaran, Ashima Arora, and Owen Rambow

4:30–5:30 As Long as You Name My Name Right: Social Circles and Social Sentiment in the Hollywood Hearings
Oren Tsur, Dan Calacci, and David Lazer

4:30–5:30 Towards Tracking Political Sentiment through Microblog Data
Yu Wang, Tom Clark, Jeffrey Staton, and Eugene Agichtein

4:30–5:30 Innovation of Verbs in Hebrew
Uzzi Ornan

4:30–5:30 User Type Classification of Tweets with Implications for Event Recognition
Lalindra De Silva and Ellen Riloff

4:30–5:30 Collective Stance Classification of Posts in Online Debate Forums
Dhanya Sridhar, Lise Getoor, and Marilyn Walker

5:30–5:40 **Closing Remarks**

Workshop 13:
Joint 1-day workshop between SIGMORPHON and SIGFSM

Organizers: *Ozlem Cetinoglu, Jeffrey Heinz, Andreas Maletti, and Jason Riggle*

Venue: Salon I

Keynote session

9:00–9:15 **Welcome by Jason Riggle**

9:15–10:15 **Keynote by Jason Eisner**

10:30–11:00 **Morning break**

Morning session

11:00–11:30 Revisiting Word Neighborhoods for Speech Recognition
Preethi Jyothi and Karen Livescu

11:30–12:00 The Error-driven Ranking Model of the Acquisition of Phonotactics: How to Keep the Faithfulness Constraints at Bay
Giorgio Magri

12:00–12:30 Comparing Models of Phonotactics for Word Segmentation
Natalie Schrimpf and Gaja Jarosz

12:30–2:00 **Lunch break**

Early afternoon session

2:00–2:30 Generalizing Inflection Tables into Paradigms with Finite State Operations
Mans Hulden

2:30–3:00 Automatic Conversion of Dialectal Tamil Text to Standard Written Tamil Text using FSTs
Marimuthu K and Sobha Lalitha Devi

3:00–3:30 Rule Based Morphological Analyzer of Kazakh Language
Gulshat Kessikbayeva and Ilyas Cicekli

3:30–4:00 **Afternoon break**

Afternoon session

4:00–4:30 Rules, Analogy, and Social Factors Codetermine Past-tense Formation Patterns in English
Peter Racz, Clayton Beckner, Jennifer B. Hay, and Janet B. Pierrehumbert

Panel discussion

4:30–4:40 10 Open Questions in Computational Morphology
Grzegorz Kondrak

4:40–5:30 **Panel discussion by Jason Eisner, Mans Hulden, Grzegorz Kondrak, Jason Riggle**

Workshop 14:
2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation

Organizers: *Teruko Mitamura, Eduard Hovy, and Martha Palmer*

Venue: Salon II

Friday, June 27, 2014

9:00–9:15 **Welcome**

9:15–10:30 **Invited Talk: Implicit Causal Relations among Events in Text by Jerry Hobbs, ISI/USC**

10:30–11:00 **Break**

11:00–12:00 **Session I: The Nature of Events**

12:00–1:00 **Lunch Break**

1:00–2:30 **Poster Session**
- Augmenting FrameNet Via PPDB
  Pushpendre Rastogi and Benjamin Van Durme
- Verbal Valency Frame Detection and Selection in Czech and English
  Ondřej Dušek, Jan Hajič, and Zdeňka Urešová
- Challenges of Adding Causation to Richer Event Descriptions
  Rei Ikuta, William Styler, Mariah Hamang, Tim O'Gorman, and Martha Palmer
- Inter-annotator Agreement for ERE annotation
  Seth Kulick, Ann Bies, and Justin Mott
- Unsupervised Techniques for Extracting and Clustering Complex Events in News
  Delia Rusu, James Hodson, and Anthony Kimball
- Conceptual and Practical Steps in Event Coreference Analysis of Large-scale Data
  Fatemeh Torabi Asr, Jonathan Sonntag, Yulia Grishina, and Manfred Stede
- A Comparison of the Events and Relations Across ACE, ERE, TAC-KBP, and FrameNet Annotation Standards
  Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis
- Is the Stanford Dependency Representation Semantic?
  Rachel Rudinger and Benjamin Van Durme
- Qualities of Eventiveness
  Sean Monahan and Mary Brunson
- Evaluation for Partial Event Coreference
  Jun Araki, Eduard Hovy, and Teruko Mitamura

2:30–3:30 **Session II: Event Ontology**

3:30–4:00 **Break**

4:00–5:00 **Session III: Event Structure and Subevents**

5:00–6:00 **Session IV: Shared Task Presentations and Discussions**

6:00–6:00 **Close**

Workshop 15:
Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929-–2014)

Organizers: *Miriam R.L. Petruck and Gerard de Melo*

Venue: Heron (4th floor)

Friday, June 27, 2014

7:30–9:00	**Registration and breakfast**
9:00–9:45	FrameNet: A Knowledge Base for Natural Language Processing *Collin Baker*
9:45–10:30	The Case for Empiricism (With and Without Statistics) *Kenneth Church*
10:30–11:00	**Break**
11:00–11:45	SemLink+: FrameNet, VerbNet and Event Ontologies *Martha Palmer, Claire Bonial, and Diana McCarthy*
11:45–12:30	Case, Constructions, FrameNet, and the Deep Lexicon *Jerry Hobbs*
12:30–2:00	**Lunch break**
2:00–2:45	Statistical Models for Frame-Semantic Parsing *Dipanjan Das*
2:45–3:30	Bridging Text and Knowledge with Frames *Srini Narayanan*
3:30–4:00	**Break**
4:00–4:45	Using Frame Semantics in Natural Language Processing *Apoorv Agarwal, Daniel Bauer, and Owen Rambow*
4:45–5:30	Who Evoked that Frame? Some Thoughts on Context Effects and Event Types *Katrin Erk*

Workshop 16:
Workshop on Interactive Language Learning, Visualizations, and Interfaces

Organizers: *Jason Chuang, Spence Green, Marti Hearst, Jeffrey Heer, and Philipp Koehn*

Venue: Galena (4th floor)

Friday, June 27, 2014

8:30–8:45 **Opening Remarks**

8:45–9:30 **Invited Talk**

9:30–10:30 **Research Papers**

9:30–9:50 MiTextExplorer: Linked brushing and mutual information for exploratory text data analysis
Brendan O'Connor

9:50–10:10 Interactive Learning of Spatial Knowledge for Text to 3D Scene Generation
Angel Chang, Manolis Savva, and Christopher Manning

10:10–10:30 Dynamic Wordclouds and Vennclouds for Exploratory Data Analysis
Glen Coppersmith and Erin Kelly

10:30–11:00 **Coffee Break**

11:00–11:45 **Invited Talk**

11:45–12:30 **Invited Talk**

12:30–2:00 **Lunch Break**

2:00–2:45 **Invited Talk**

2:45–4:00 **Interactive Demo Session**
- Active Learning with Constrained Topic Model
  Yi Yang, Shimei Pan, Doug Downey, and Kunpeng Zhang
- GLANCE Visualizes Lexical Phenomena for Language Learning
  MeiHua Chen, Shih-Ting Huang, Ting-Hui Kao, Hsun-wen Chiu, and Tzu-Hsi Yen
- SPIED: Stanford Pattern based Information Extraction and Diagnostics
  Sonal Gupta and Christopher Manning
- Interactive Exploration of Asynchronous Conversations: Applying a User-centered Approach to Design a Visual Text Analytic System
  Enamul Hoque, Giuseppe Carenini, and Shafiq Joty
- MUCK: A toolkit for extracting and visualizing semantic dimensions of large text collections
  Rebecca Weiss
- Design of an Active Learning System with Human Correction for Content Analysis
  Nancy McCracken, Jasy Suet Yan Liew, and Kevin Crowston
- LDAvis: A method for visualizing and interpreting topics
  Carson Sievert and Kenneth Shirley

- Hiearchie: Visualization for Hierarchical Topic Models
  Alison Smith, Timothy Hawes, and Meredith Myers
- Concurrent Visualization of Relationships between Words and Topics in Topic Models
  Alison Smith, Jason Chuang, Yuening Hu, Jordan Boyd-Graber, and Leah Findlater

4:00–4:45 **Invited Talk**

4:45–5:30 **Discussion and Closing Remarks**

Workshop 17:
Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality

Organizers: *Philip Resnik, Rebecca Resnik, and Margaret Mitchell*

Venue: Salon VII

Friday, June 27, 2014

9:00–9:10 **Introduction**

9:10–10:30 **Presentations and discussion 1**

- Predicting military and veteran suicide risk: Cultural aspects
  Paul Thompson, Craig Bryan, and Chris Poulin
- Linguistic Indicators of Severity and Progress in Online Text-based Therapy for Depression
  Christine Howes, Matthew Purver, and Rose McCabe

10:30–11:00 **Morning break**

11:00–12:20 **Presentations and discussion 2**

- Comparison of different feature sets for identification of variants in progressive aphasia
  Kathleen C. Fraser, Graeme Hirst, Naida L. Graham, Jed A. Meltzer, Sandra E. Black, and Elizabeth Rochon
- Aided diagnosis of dementia type through computer-based analysis of spontaneous speech
  William Jarrold, Bart Peintner, David Wilkins, Dimitra Vergryi, Colleen Richey, Maria Luisa Gorno-Tempini, and Jennifer Ogar

12:20–12:30 **Poster teasers**

12:30–2:00 **Lunch (provided) and poster session**

2:00–3:30 **Presentations and discussion 3**

- Assessing Violence Risk in Threatening Communications
  Kimberly Glasgow and Ronald Schouten
- Detecting linguistic idiosyncratic interests in autism using distributional semantic models
  Masoud Rouhizadeh, Emily Prud'hommeaux, Jan van Santen, and Richard Sproat

3:30–4:00 **Afternoon break**

4:00–4:45 **Presentations and discussion 4**

- Quantifying Mental Health Signals in Twitter
  Glen Coppersmith, Mark Dredze, and Craig Harman

4:45–5:30 **General discussion**

Posters

- Applying prosodic speech features in mental health care: An exploratory study in a life-review intervention for depression
  Sanne M.A. Lamers, Khiet P. Truong, Bas Steunenberg, Franciska de Jong, and Gerben J. Westerhof
- Challenges in Automating Maze Detection
  Eric Morley, Anna Eva Hallin, and Brian Roark
- Learning Predictive Linguistic Features for Alzheimer's Disease and related Dementias using Verbal Utterances
  Sylvester Olubolu Orimaye, Jojo Sze-Meng Wong, and Karen Jennifer Golden
- Linguistic and Acoustic Features for Automatic Identification of Autism Spectrum Disorders in Children's Narrative
  Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, and Satoshi Nakamura
- Mining Themes and Interests in the Asperger's and Autism Community
  Yangfeng Ji, Hwajung Hong, Rosa Arriaga, Agata Rozga, Gregory Abowd, and Jacob Eisenstein
- Toward Macro-Insights for Suicide Prevention: Analyzing Fine-Grained Distress at Scale
  Christopher Homan, Ravdeep Johar, Tong Liu, Megan Lytle, Vincent Silenzio, and Cecilia Ovesdotter Alm
- Towards Assessing Changes in Degree of Depression through Facebook
  H. Andrew Schwartz, Johannes Eichstaedt, Margaret L. Kern, Gregory Park, Maarten Sap, David Stillwell, Michal Kosinski, and Lyle Ungar

8
Anti-harassment policy

ACL has always prided itself as a venue that allows for the open exchange of ideas and the freedom of thought and expression. In keeping with these beliefs, to codify them and to ensure that ACL becomes immediately aware of any deviation from these principles, ACL has instituted an anti-harassment policy in coordination with NAACL.

Harassment and hostile behavior is unwelcome at any ACL conference; including speech or behavior that intimidates, creates discomfort, or interferes with a person's participation or opportunity for participation in the conference. We aim for ACL conferences to be environments 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.

If you are a victim of harassment or hostile behaviour at an ACL conference, or otherwise observe such behaviour toward someone else, please contact any of the following people:

- Any current member of the ACL board (`http://aclweb.org/about`)

- Hal Daume III

- Julia Hirschberg

- Su Jian

- Priscilla Rasmussen

Please be assured that if you approach us, your concerns will be kept in strict confidence, and we will consult with you on the actions taken by the Board.

The full policy and its implementation is defined at:

`http://naacl.org/policies/anti-harassment.html`

Index

Food and drink

ACL 2014 is located in Baltimore's newest upscale neighborhood, Harbor East, which is nestled between touristy Inner Harbor to the west and historic Fells Point to the east. Head west along the waterfront to the Inner Harbor for the Aquarium, Science Center, and dozens of chain restaurants. Stay in Harbor East for good but often somewhat expensive restaurants and bars. Head further east along the waterfront past Caroline Street to Fells Point for cobblestone streets and centuries old bars and restaurants. Baltimore's Little Italy is just a couple blocks north from the conference hotel.

Restaurants, Bars, and Cafes by Neighborhood

Throughout this guide, we provide distances away from the Marriott Waterfront conference hotel. For restaurants serving full meals, we also provide approximate relative food costs on a four point ($ to $$$$) scale. We highlight some of our very favorite spots as *editor favorites*.

Harbor East   The conference hotel is located in Harbor East, a fast developing, upscale neighborhood that replaced decaying industrial warehouses about a decade ago. In addition to a few nice restaurants, the neighborhood offers plenty of shopping, a movie theater, and easy access to Baltimore's free circulator bus and water taxi.

- Lebanese Taverna (719 S. President, 0.1 miles away): upscale Lebanese food; good for lunch or dinner and plenty of space for large groups. $$$.

- Cinghiale (822 Lancaster, 0.1 miles away): upscale Italian fare and extensive wine list. $$$$.

- Whole Foods (1001 Fleet, 0.2 miles away); predictable grocery store fare; hot and cold food bars for a quick meal. $.

- Gordon Biersch (1000 Lancaster, 0.2 miles away): new, large bar, good for watching sports or a quick beer near the conference hotel. $$.

- Teavolve (1401 Aliceanna, 0.3 miles away): great for working (free wifi), morning coffee, and lunch. $.

- RA Sushi Bar Restaurant (1390 Lancaster, 0.3 miles away): Good lunch specials, great for large groups. $$.

- Fleet Street Kitchen (1012 Fleet, 0.3 miles away): Upscale, beautiful interior serving excellent local meat and produce. $$$. *Editor favorite.*

Little Italy　Baltimore's historic Italian neighborhood is small but bustling with delis, restaurants, and cafes. Several old warehouses nearby have also been converted into retail and office space. Take a walk up Albemarle and High Streets after dark, when they are lit with charming street lights.

- Vaccaro's (222 Albemarle, 0.3 miles away): Italian bakery; stop by for a late night coffee and dessert. $.

- Chiapperelli's (237 S. High, 0.3 miles away): Our favorite restaurant in Little Italy. Good for lunch or dinner, and plenty of space for big groups. $$. *Editor favorite.*

- Heavy Seas Alehouse (1300 Bank, 0.4 miles away): Local brewpub (the brewery itself a few miles away); upscale bar food and outdoor beer garden. $$$.

- Amicci's (231 S. High, 0.3 miles away). Classic, reasonably priced Italian fare. $$.

- La Tavola (248 Albemarle, 0.3 miles away). One of the nicer, but pricier options in the neighbhorhood. $$$.

- Ciao Bella (236 S High, 0.3 miles away): You can probably find an online coupon, but the food and service are not as high quality as other spots in the neighborhood. $$.

- Mo's Fisherman's Wharf (219 S. President, 0.3 miles away): Tourist trappy and not particularly nice, but an easy walk from the conference center. $$.

- My Thai (323 S.Central, 0.4 miles away): Solid Thai food, good for big groups. $$.

Fells Point　Fells Point is a quintessential Baltimore neighborhood. Stroll along cobblestoned Thames Street for a picturesque view of the harbor, check out the centuries old rowhouses on Lancaster and Ann Streets, and dine out in one of the plethora of neighborhood restaurants. Locals claim that Fells Point has the highest number of bars per capita of any neighborhood in the United States! Be warned that many restaurants are small and won't be able to accommodate very large groups.

- One-Eyed Mike's (708 S Bond, 0.4 miles away): Small, cozy spot for a meal or drink. $$.

- Blue Moon Cafe (1621 Aliceanna, 0.5 miles away): Legendary-ish all-day (and -night) breakfast place. Long lines, delicious food. $.

- Brick Oven Pizza (800 S. Broadway, 0.5 miles away): Pizza by the slice, BYOB, eat-in. $.

- Bertha's (734 S. Broadway, 0.5 miles away): Touristy seafood restaurant with surprisingly great food. Eat Bertha's mussels. $$.

- Sticky Rice (1634 Aliceanna, 0.5 miles away): Hipster Sushi. Come early for a great happy hour. $$.

- Birds of a Feather (1712 Aliceanna, 0.6 miles away): Quiet, local scotch bar. $$.

- Max's Taphouse (737 S Broadway, 0.6 miles away): Bro taphouse. Over 100 beers on tap. $$.

- The Horse You Came In On Saloon (1626 Thames, 0.6 miles away): Lively bar with live music everyday. Edgar Allan Poe had his last drink here before he was found dying in a nearby ditch. $$.

- Bond Street Social (901 S Bond, 0.6 miles away): Upscale bar/restaurant with great outdoor seating. $$$.

- Kooper's Tavern (1702 Thames, 0.6 miles away): A favorite of locals and tourists alike! Bar/restaurant with excellent burgers and beer selection. $$. *Editor favorite.*

- Duda's Tavern (1600 Thames, 0.6 miles away): Small bar/restaurant, local hangout, great crabcakes and friendly service. $$. *Editor favorite.*

- Mezze (1606 Thames, 0.6 miles away): Good tapas with some outdoor seating. Make a reservation for large groups. $$$.

- Bar (Lancaster, 0.6 miles away): Local favorite. The service is bad, beer is cheap, food is unavailable, and pool table is broken, but Bar is always a fun night out. $. *Editor favorite.*

- Thames Street Oyster House (1728 Thames, 0.7 miles away): Delicious but pricey seafood. $$$.

- Wharf Rat (801 S. Ann, 0.7 miles away): Exceptionally cozy bar that serves not-great bar food but has an excellent beer selection and a pool table in the back. $.

- Nanami (907 S. Ann, 0.7 miles away): Small, cozy sushi on the water. $$.

- Tortilleria Sinaloa (1716 Eastern, 0.7 miles away): There's a lot of good Mexican food on Eastern Ave., but this is one of the best. Stop by for a tamale for lunch. $. *Editor favorite.*

- Darbar (1911 Aliceanna, 0.7 miles away): Fells Point's only Indian restaurant. Good for large groups at lunch or dinner. $$.

- John Steven, Ltd (1800 Thames, 0.7 miles away); Small bar with decent food. Just on the edge of Fells Point rowdiness. $$.

- Daily Grind (1720 Thames, 0.7 miles away): Morning coffee or a quick lunch; free wifi and plenty of space to work. $.

- Ale Mary's (1939 Fleet, 0.8 miles away): Worth the walk! Small bar/restaurant. Try the tater tots and Chicken Chesapeake. $$. *Editor favorite.*

- Red Star Bar and Grill (906 S Wolfe, 0.8 miles away): Very good bar food and nice, historic building. Worth the extra walk. $$.

- Spirits Tavern (1901 Bank, 0.9 miles away): Located in a former funeral home, Spirits is a real neighborhood gem. Come to meet some friendly locals. $. *Editor favorite.*

Inner Harbor   Baltimore's Inner Harbor was revitalized in the 1970s and 80s and has since served as the city's tourist hub. However, it has declined some in recent years as new development has preferred other waterfront neighborhoods, including Harbor East. The neighborhood does still offer an excellent waterfront view of the city and the habitual array of chain restaurants. We recommend checking out Yelp or TripAdvisor for details on the restaurants listed below, as the editors plead the 5th on opinion-giving.

- McCormick & Schmick's (711 Eastern, 0.2 miles away): Expensive dining very near the conference hotel. $$$.

- Hard Rock Cafe (601 E Pratt, 0.4 miles away): Wait, this restaurant didn't go out of business in the 90s? $$.

- Blu Bambu (621 E Pratt, 0.4 miles away): Mediocre Asian fast food. $$.

- Phillips Seafood (601 E Pratt, 0.5 miles away): Baltimore's famous seafood restaurant. $$$.

- Dick's Last Resort (621 E Pratt, 0.5 miles away): [Read like a letter of recommendation:] Aptly named. Nice view of the Inner Harbor? $$.

- Fogo de Chao (600 E Pratt, 0.5 miles away): Meat-on-a-stick, Brazilian-style steakhouse chain. $$$.

- Chipotle (621 E Pratt, 0.5 miles away): Quick, cheap burritos. $.

- Five Guys (201 E Pratt, 0.7 miles away): Good fast food burgers. $.

- The Capital Grille (500 E Pratt, 0.5 miles away): Good, expensive steaks (so they say). $$$$.

- Ruth's Chris Steak House (600 Water, 0.6 miles away): Good, expensive steaks (so they say). $$$$.

- Ruth's Chris Steak House (711 Eastern, 0.4 miles away): Sample size of 2 on this one. Remember your significance testing! $$$$.

- Cheesecake Factory (201 E Pratt, 0.7 miles away): You can order more than just Cheesecake, but you'd be hard pressed to find an entree with fewer than 2,000 calories. $$.

- Bubba Gump Shrimp Co. (301 Light, 0.9 miles away): Life is like a box of so-so seafood. $$.

- Rusty Scupper. (402 Key Hwy, 1.6 miles away): Expensive seafood dining across the inner harbor from the conference hotel. $$$.

Mount Vernon & Downtown   Just north of the Inner Harbor, the Mount Vernon neighborhood originally was inhabited by Baltimore's most wealthy and fashionable families (and today, graduate students). It features lovely brownstone architecture, the original Washington Monument, the Peabody Conservatory, the Walters Art Museum, as well as a great selection of bars and restaurants. Station North, just north of Mount Vernon, is home to some more artsy and divy venues.

- Iggie's (818 N. Calvert, 1.5 miles away): Great thin-crust pizza, unconventional menu, BYOB. $$. *Editor favorite.*

- Stang of Siam (1301 Calvert, 1.8 miles away): Lovely Thai restaurant with good food. Understanding of spiciness randomized daily. $$.

- Helmand (806 N Charles, 1.8 miles away): Afghan cuisine, often listed amongst best restaurants in Baltimore. Owned by the brother of Afghan president Hamid Karzai. $$$.

- Brewer's Art (1106 N Charles, 1.9 miles away): Great brewpub with a fantastic selection of beers. Brewer's was voted America's Best Bar 2008, and is the largest brewery in Baltimore. It features an upscale restaurant section as well as a divy basement. $$. *Editor favorite.*

- Red Maple (930 N Charles, 2.1 miles away): Likely Baltimore's second-best night club. Possibly Baltimore's worst night club. Allegedly serves food. $$.

- Club Charles (1724 N Charles, 2.3 miles away): Divy bar lit in gloomy pink. Rather mediocre drinks, but occasional trapeze performances. $.

- Joe Squared (133 W North, 2.5 miles away): Pizza and beer place, hosts local live music of varying quality on most nights. Build your own pizza from a wealth of delicious, partially roof-grown toppings. $$.

- Alewife (21 N Eutaw, 1.5 miles away): Located near the Hippodrome and Lexington Market, this restaurant-bar sports an ever-changing draft list of 40 well-selected craft beers, along with a near-interminable list of bottled beers. Their burger is one of the best in the city. $$. *Editor favorite*.

Federal Hill   Federal Hill is another popular Baltimore neighborhood for eating, drinking, and residing. Located near both the Orioles and Ravens stadiums, folks in this part of town are even more fanatical about our local teams than most Baltimore residents. From the conference hotel, you can reach it by taking the free water taxi straight across the water to the south. Federal Hill Park provides excellent views of the city, and the neighborhood is a close second in the number of bars per capita ranking.

- Little Havana (1325 Key Highway, 2.5 miles away on land, 0.7 miles on water): Take the water taxi across the harbor and enjoy cuban cuisine on the water. Good for large groups. $$.

- Thai Arroy (1019 Light, 1.7 miles away on land, 0.9 miles away on water): Great, no-frills Thai food (and it's BYOB!). This restaurant is small and won't be able to easily accommodate large groups. $$.

- Cross Street Market (1065 S Charles, 1.8 miles away on land, 1.0 miles on water): Classic Baltimore city market. Buy some fresh flowers, produce, and a burner phone before grabbing a beer, crabcakes, oysters, shrimp, and sushi from Nick's, on the west end of the market. Well-worth the trip around or across the harbor. $. *Editor favorite.*

- Byblos (1033 Light, 1.7 miles away on land, 0.9 miles on water): Inexpensive Lebanese food with the sweetest owners we've ever met. (And it's BYOB!). $.

- Mother's Federal Hill Grille (1113 S Charles, 1.8 miles away on land, 1.0 miles on water): One of Baltimore's staple bar/restaurants. Mother's is huge but often full. It's a great place to celebrate local sports teams' wins, linger over weekend brunch, or grab a quick lunch or happy hour beer. $$.

Hampden Formerly a blue collar neighborhood settled to provide housing for mills workers, Hampden is now one of Baltimore's premier hipster and artist neighborhoods. Plenty of small, kooky stores, good food, and local color are to be found around W 36th street, known as "The Avenue." Hampden is a bit of a trek from the conference venue, but well worth an afternoon stroll if you have the time.

- Grano Pasta Bar (1031 W 36th, 4.2 miles away): Tiny Grano is an amazing, family-run no-frills pasta bar. Freshly made pasta and delicious sauces. BYOB. Big Grano is Tiny Grano's big brother, in fully-grown Italian restaurant form. $. *Editor favorite.*

- Luigi's Italian Deli (846 W 36th, 4.3 miles away): Delicious, richly topped Italian sandwiches. $.

- De Kleine Duivel (3602 Hickory, 4.2 miles away): Simple bar specializing in Belgian beers. Good selection and plenty of space for groups. $.

- Corner BYOB (850 W 36th, 4.4 miles away): Upscale (and cash-only) restaurant that features an adventurous eaters club which serves (amongst other things) python meat. Says BYOB on every available surface in the restaurant. May be BYOB. $$$.

- Spro (851 W 36th, 4.4 miles away): Hampden's solid take on the fancy coffee shop. $.

- Woodberry Kitchen (2010 Clipper Park Road, 4.8 miles away): Excellent restaurant by Hampden. Local, seasonal New American cuisine. Great for dinner and weekend brunch, great cocktails. From the Inner Harbor, take the light rail North to Woodberry station, or walk down from Hampden. Reservations highly recommended. $$$. *Editor favorite.*

- Artifact Coffee (1500 Union, 4.5 miles away): Woodberry Kitchen's coffee shop cousin. Great coffee, pastries, and small plates. Well worth a drop-in if you're in the area. $$.

Foodie Eats Baltimore is home to a number of great restaurants, serving anything from local New American fare to Italian or Afghani cuisine. Though not all are nearby, these places are certainly worth a dinner trip.

- Woodberry Kitchen (2010 Clipper Park Road, 4.8 miles away): Excellent restaurant by Hampden. Local, seasonal New American cuisine. Great for dinner and weekend brunch, fantastic cocktails. From the Inner Harbor, take the light rail North to Woodberry station, or walk down from Hampden. Reservations highly recommended. $$$. *Editor favorite.*

- Waterfront Kitchen (1417 Thames, 0.5 miles away): Situated right on the water of the harbor, this restaurant serves American cuisine with an emphasis on local ingredients. $$$.

- Helmand (806 N Charles, 1.8 miles away): Afghan cuisine, often listed amongst best restaurants in Baltimore. Owned by the brother of Afghan president Hamid Karzai. $$$.

- Salt (2127 E Pratt, 1.3 miles away): Upscale New American restaurant in Upper Fells Point. $$$.

- Cinghiale (822 Lancaster, 0.1 miles away): Upscale Italian fare and extensive wine list. $$$$.

Fancy [Expensive] Eats

- McCormick & Schmick's (711 Eastern, 0.2 miles away): Expensive dining very near the conference hotel. $$$.

- Ruth's Chris Steak House (600 Water, 0.6 miles away): Good, expensive steaks (so they say). $$$$.

- Black Olive (814 Bond, 0.4 miles away): Mediterranean cuisine, with an emphasis on seafood. $$$.

- The Prime Rib (1101 N Calvert, 1.7 miles away): The best steakhouse in Baltimore, according to some. $$$$.

Transportation

From BWI to Conference Hotel   The cheapest is to take the Light Rail Camden Line to Camden Yards and then walk, taxi, or take the circulator bus one mile east through downtown to the conference hotel. A slightly more expensive option is the SuperShuttle, a shared van service that costs about $15 per person. It should cost about $35 to reach the conference hotel in a private taxi. Note that shuttle and taxi drivers should also be tipped 10-15%.

From Penn Station to Conference Hotel   The cost and time involved in traveling from Baltimore's main train station to the conference hotel will vary depending on traffic. The trip is about 3 miles and should cost no more than $15 in light traffic. However, during both morning and evening rush hour, the trip could be frustratingly slow. Be warned!

City Public Transportation

- Circulator: Baltimore city offers a free bus service, the Charm City Circulator. Several routes connect the neighborhoods surrounding the harbor with one another and Mount Vernon. More information can be found at `http://www.charmcitycirculator.com/`. The next stop and bus ETA can be found on `www.nextbus.com`.

- Water Taxi: the Baltimore Water Taxi connects a number of spots along the harbor by boat. Its service includes the free Harbor Connector that connects the conference venue in Harbor East with Federal Hill on the other side of the harbor.

To/From DC   Take the free Orange line Circulator bus (which picks up just around the corner from the conference hotel) to Camden Yards and then hop on the Camden Line MARC train to DC's Union Station. Another option is to take a taxi to Penn Station and then Amtrak or the Penn MARC train to DC's Union Station. Train service to and from Baltimore's Penn Station is much more frequent than the Camden line.

Cultural Attractions

Art Museums

- American Visionary Art Museum (800 Key Highway, 1.8 miles away, or 0.8 miles via the ferry): Located in a former whiskey warehouse, Baltimore's AVAM offers an avant-garde art collection that is worth experiencing.*Editor's favorite.*

- Walter's Art Museum (600 N Charles, 1.6 miles away): Building on the 1931 donation by one of Baltimore's richest industrialists, this museum offers a collection of art and artifacts spanning a wide range of cultures and eras. Free admission.

- Baltimore Museum of Art (10 Art Museum Dr, 4.0 miles away): The BMA houses several world-class collections of art, both classical and contemporary. Located right on the Johns Hopkins University Campus, it also sports a beautiful sculpture garden. Free admission.

History

- Fort McHenry (2400 E Fort): See the fort where Francis Scott Key wrote the Star Spangled Banner 200 years ago this September! The Fort is planning a big celebration of the end of the war of 1812 and has upgraded its grounds substantially in recent years. Be sure to leave time for a walk around the one mile waterfront path.

- Reginald F. Lewis Museum of Maryland African American History & Culture (830 E Pratt, 0.3 miles away): Located very close to the conference hotel, the Reginald F. Lewis Museum is an excellent museum highlighting the accomplishments of African Americans, with a focus on those from Maryland. Admission is reasonably priced ($6 for students with a valid ID, $8 for adults).

- Baltimore's Historic Ships: See a variety of historic ships, including the US Constellation and a submarine. The ships are located throughout the Inner Harbor. Visit `http://www.historicships.org/` for more information.

Others

- Chessie Paddle Boating (Inner Harbor Pier 1, 0.8 miles away): Rent a dragon paddle boat and drive yourself around the Inner Harbor waters.

- Maryland Science Center (601 Light, 1.1 miles away): Very nice museum, particularly for kids. It's in the Inner Harbor at the intersection of Key Highway and Light Street.

- Urban Pirates Cruise (913 S Ann, 0.7 miles away): Family-friendly cruises by day adult-fun cruises by night, and friendly, energetic crews all day long. Leaves from the end of Ann Street in Fells Point, just next to Nanami restaurant.

- Spirit Dinner Cruises from the Inner Harbor (561 Light St, 0.9 miles away): Overpriced dinner provides a nice view of the city from the water.

- Orioles Major League Baseball Game (333 W Camden, 1.2 miles away): The Orioles will play the Chicago White Sox on Monday, Tuesday, Wednesday during ACL and the Rays on Friday and Saturday. Take the orange circulator bus from President Street Circle to Eutaw Street to see MLB's most beautiful stadium. Tickets are always available day-of from the box office. Go O's!

- Live professional theater production of the acclaimed satirical play about language *The Memo* by Václav Havel - 8PM on Thursday June 26 in Arellano Theater of Johns Hopkins University. This special reprise performance of a recent sold-out production of Baltimore's award-winning Single Carrot Theater will be preceded by a 7:15PM welcome reception and optional tour of Johns Hopkins University and its Center for Language and Speech Processing. See page 118 for more detail. Tickets are *free* at the ACL registration desk to the first 100 requesting attendees.

Exercise

There's no need to let your fitness suffer at ACL 2014! Go for a run, play beach volleyball, or check out a local crossfit or yoga spot.

Running Routes   The 7 mile Harbor Promenade provides a great, flat, scenic, and easily navigable running route. It goes right by the conference hotel.

- Westward: Head west along the promenade and turn around at the Aquarium (1 mile out and back), Maryland Science Center (2 miles out and back), the Harborview Water Taxi stop (3 miles out and back), or continue along Key Highway to the Under Armour headquarters (5.5 miles out and back) or to Hull Street, Fort Avenue, and Ft McHenry (7.5 miles to the Fort and back, 8.5 including a loop around scenic Ft McHenry).

- Eastward: Run east along the promenade and turn around at Caroline (1 mile out and back), Henderson's Wharf (2 miles out and back), Captain James Landing (3 miles out and back), the Boston Street Pier Park/Safeway (4 miles out and back), or the Canton Waterfront Park (5 miles out and back).

- Northward: Please don't run very far north.

- Hills!: If you're looking for a hillier route, go east on Eastern Avenue to Patterson Park, a large city park with lots of recreation opportunities, including an outdoor public swimming pool and tennis courts.

Crossfit   Out-of-town crossfitters are welcome to drop in at CrossFit Harbor East, located at 510 S. Eden Street. Fill out the waiver online before coming to class to save time.

Swimming   The Inner Harbor is the head of the Patapsco River. Go for a swim just outside of the hotel (just kidding, don't do this).

Yoga   The Fells Point Charm City Yoga studio is located at 1807 Thames and Sanctuary Bodyworks is at 701 South Ann. Both studios welcome drop-ins; check out their schedules online!

Beach Volleyball   Play volleyball on sand courts located right by Baltimore's beautiful waterfront. Check out `www.baltimorebeach.com` for more information.

Elite Fitness tours   Elite Fitness Tours offers a running tour through some of Baltimore's most historic neighborhoods, including Federal Hill and Mt. Vernon. Check out their website for tour information.

speech@amazon

Amazon's speech and language team's mission is to push the envelope in automatic speech recognition (ASR) and natural language understanding (NLU) in order to provide the best possible experience for our customers. We are a rapidly growing team and are looking for passionate, entrepreneurial scientists who have deep expertise in ASR, NLU, and machine learning. Please visit us at our booth or send us an email to `speech@amazon.com` for more information. For a list of current global opportunities, please visit: `www.amazon.com/speech`
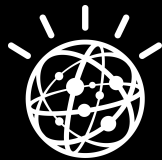
Research at Google

Google is proud to sponsor the 2014 ACL conference. Visit our booth to find out more about the latest opportunities and development from Research at Google. Follow our Google+ and Twitter pages for announcements of the research behind products, deep dives on publications, Googler interviews, collaboration stories and more!

Food and Drink Quick Reference

Establishment	Neighborhood	Address	Distance	Rating
Lebanese Taverna	Harbor East	719 S. President	0.1 miles	$$$
Cinghiale	Harbor East	822 Lancaster	0.1 miles	$$$$
Whole Foods	Harbor East	1001 Fleet	0.2 miles	$
Gordon Biersch	Harbor East	1000 Lancaster	0.2 miles	$$
McCormick & Schmick's	Inner Harbor	711 Eastern	0.2 miles	$$$
Teavolve	Harbor East	1401 Aliceanna	0.3 miles	$
RA Sushi Bar Restaurant	Harbor East	1390 Lancaster	0.3 miles	$$
★ Fleet Street Kitchen	Harbor East	1012 Fleet	0.3 miles	$$$$
Vaccaro's	Little Italy	222 Albemarle	0.3 miles	N/A
★ Chiapperelli's	Little Italy	237 S. High	0.3 miles	$$
Heavy Seas Alehouse	Little Italy	1300 Bank	0.4 miles	$$$
My Thai	Little Italy	323 S.Central	0.4 miles	$$
One-Eyed Mike's	Fells Point	708 S Bond	0.4 miles	$$
Hard Rock Cafe	Inner Harbor	601 E Pratt	0.4 miles	$$
Blu Bambu	Inner Harbor	621 E Pratt	0.4 miles	$$
Ruth's Chris Steak House	Inner Harbor	711 Eastern	0.4 miles	$$$$
Blue Moon Cafe	Fells Point	1621 Aliceanna	0.5 miles	$
Brick Oven Pizza	Fells Point	800 S. Broadway	0.5 miles	$
Bertha's	Fells Point	734 S. Broadway	0.5 miles	$$
Sticky Rice	Fells Point	1634 Aliceanna	0.5 miles	$$
Phillips Seafood	Inner Harbor	601 E Pratt	0.5 miles	$$$
Dick's Last Resort	Inner Harbor	621 E Pratt	0.5 miles	$$
Fogo de Chao	Inner Harbor	600 E Pratt	0.5 miles	$$$
Chipotle	Inner Harbor	621 E Pratt	0.5 miles	$
The Capital Grille	Inner Harbor	500 E Pratt	0.5 miles	$$$$
Birds of a Feather	Fells Point	1712 Aliceanna	0.6 miles	N/A
Max's Taphouse	Fells Point	737 S Broadway	0.6 miles	N/A
The Horse You Came In On Saloon	Fells Point	1626 Thames	0.6 miles	N/A
Bond Street Social	Fells Point	901 S Bond	0.6 miles	$$$
★ Kooper's Tavern	Fells Point	1702 Thames	0.6 miles	$$
★ Duda's Tavern	Fells Point	1600 Thames	0.6 miles	$$
Mezze	Fells Point	1606 Thames	0.6 miles	$$$
Bar	Fells Point	Lancaster	0.6 miles	N/A
Ruth's Chris Steak House	Inner Harbor	600 Water	0.6 miles	$$$$
Thames Street Oyster House	Fells Point	1728 Thames	0.7 miles	$$$
Wharf Rat	Fells Point	801 S. Ann	0.7 miles	$
Nanami	Fells Point	907 S. Ann	0.7 miles	$$
★ Tortilleria Sinaloa	Fells Point	1716 Eastern	0.7 miles	$
Darbar	Fells Point	1911 Aliceanna	0.7 miles	$$
John Steven, Ltd	Fells Point	1800 Thames	0.7 miles	$$
Daily Grind	Fells Point	1720 Thames	0.7 miles	$
Five Guys	Inner Harbor	201 E Pratt	0.7 miles	$
Cheesecake Factory	Inner Harbor	201 E Pratt	0.7 miles	$$
★ Ale Mary's	Fells Point	1939 Fleet	0.8 miles	$$
Red Star Bar and Grill	Fells Point	906 S Wolfe	0.8 miles	$$
★ Spirits Tavern	Fells Point	1901 Bank	0.9 miles	N/A
Bubba Gump Shrimp Co.	Inner Harbor	301 Light	0.9 miles	$$
Iggie's	Mt. Vernon	818 N Calvert	1.5 miles	$$
Thai Arroy	Federal Hill	1019 Light	1.7 miles	$$
Byblos	Federal Hill	1033 Light	1.7 miles	$
Helmand	Mt. Vernon	806 N Charles	1.8 miles	$$$
★ Cross Street Market	Federal Hill	1065 S Charles	1.8 miles	$
Mother's	Federal Hill	1113 S Charles	1.8 miles	$$
★ Brewer's Art	Mt. Vernon	1106 N Charles	1.9 miles	$$

★ = Editors' choice. For details and descriptions and additional local information, please see p.177.

ACL 2014 gratefully acknowledges the following sponsors for their support:

Platinum Level:

Gold Level:

Silver Level:

Bronze Level:

Supporters: